Network Working Group
Request for Comments: 4426
Category: Standards Track

J. Lang, Ed.
B. Rajagopalan, Ed.
D. Papadimitriou, Ed.

March 2006

Generalized Multi-Protocol Label Switching (GMPLS) Recovery Functional Specification

Status of This Memo

Copyright Notice

Abstract

This document presents a functional description of the protocol extensions needed to support Generalized Multi-Protocol Label Switching (GMPLS)-based recovery (i.e., protection and restoration). Protocol specific formats and mechanisms will be described in companion documents.

Table of Contents

1. Introduction

A requirement for the development of a common control plane for both optical and electronic switching equipment is that there must be signaling, routing, and link management mechanisms that support data plane fault recovery. In this document, the term "recovery" is generically used to denote both protection and restoration; the specific terms "protection" and "restoration" are used only when differentiation is required. The subtle distinction between protection and restoration is made based on the resource allocation done during the recovery period (see [RFC4427]).

A label-switched path (LSP) may be subject to local (span), segment, and/or end-to-end recovery. Local span protection refers to the protection of the link (and hence all the LSPs marked as required for span protection and routed over the link) between two neighboring switches. Segment protection refers to the recovery of an LSP segment (i.e., an SNC in the ITU-T terminology) between two nodes, i.e., the boundary nodes of the segment. End-to-end protection refers to the protection of an entire LSP from the ingress to the egress port. The end-to-end recovery models discussed in this document apply to segment protection where the source and destination refer to the protected segment rather than the entire LSP. Multiple recovery levels may be used concurrently by a single LSP for added resiliency; however, the interaction between levels affects any one direction of the LSP results in both directions of the LSP being switched to a new span, segment, or end-to-end path.

Unless otherwise stated, all references to "link" in this document indicate a bi-directional link (which may be realized as a pair of unidirectional links).

Consider the control plane message flow during the establishment of an LSP. This message flow proceeds from an initiating (or source) node to a terminating (or destination) node, via a sequence of intermediate nodes. A node along the LSP is said to be "upstream" from another node if the former occurs first in the sequence. The latter node is said to be "downstream" from the former node. That is, an "upstream" node is closer to the initiating node than a node further "downstream". Unless otherwise stated, all references to "upstream" and "downstream" are in terms of the control plane message flow.

The flow of the data traffic is defined from ingress (source node) to egress (destination node). Note that for bi-directional LSPs, there are two different data plane flows, one for each direction of the LSP. This document presents a protocol functional description to support Generalized Multi-Protocol Label Switching (GMPLS)-based recovery (i.e., protection and restoration). Protocol-specific formats, encoding, and mechanisms will be described in companion documents.

1.1. Conventions Used in This Document

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

In addition, the reader is assumed to be familiar with the terminology used in [RFC3945], [RFC3471] and referenced as well as [RFC4427].

2. Span Protection

Consider a (working) link i between two nodes A and B. There are two fundamental models for span protection. The first is referred to as 1+1 protection. Under this model, a dedicated link j is pre-assigned to protect link i. LSP traffic is permanently bridged onto both links i and j at the ingress node, and the egress node selects the signal (i.e., normal traffic) from i or j, based on a selection function (e.g., signal quality). Under unidirectional 1+1 span protection (Section 2.1), each node A and B acts autonomously to select the signal from the working link i or the protection link j. Under bi-directional 1+1 span protection (Section 2.2) the two nodes A and B coordinate the selection function such that they select the signal from the same link, i or j.

Under the second model, a set of N working links are protected by a set of M protection links, usually with M =< N. A failure in any of the N working links results in traffic being switched to one of the M protection links that is available. This is typically a three-step process: first the data plane failure is detected at the egress node and reported (notification), then a protection link is selected, and finally, the LSPs on the failed link are moved to the protection link. If reversion is supported, a fourth step is included, i.e., return of the traffic to the working link (when the working link has recovered from the failure). In Section 2.3, 1:1 span protection is described. In Section 2.4, M:N span protection is described, where M =< N.

2.1. Unidirectional 1+1 Dedicated Protection

Suppose a bi-directional LSP is routed over link i between two nodes A and B. Under unidirectional 1+1 protection, a dedicated link j is pre-assigned to protect the working link i. LSP traffic is permanently bridged on both links at the ingress node, and the egress node selects the normal traffic from one of the links, i or j. If a node (A or B) detects a failure of a span, it autonomously invokes a process to receive the traffic from the protection span. Thus, it is possible that node A selects the signal from link i in the B to A direction of the LSP, and node B selects the signal from link j in the A to B direction.

The following functionality is required for 1+1 unidirectional span protection:

- Routing: A single TE link encompassing both working and protection links SHOULD be announced with a Link Protection Type "Dedicated 1+1", along with the bandwidth parameters for the working link. As the resources are consumed/released, the bandwidth parameters of the TE link are adjusted accordingly. Encoding of the Link Protection Type and bandwidth parameters in IS-IS is specified in [RFC4205]. Encoding of this information in OSPF is specified in [RFC4203].

- Signaling: The Link Protection object/TLV SHOULD be used to request "Dedicated 1+1" link protection for that LSP. This object/TLV is defined in [RFC3471]. If the Link Protection object/TLV is not used, link selection is a matter of local policy. No additional signaling is required when a fail-over occurs.

- Link management: Both nodes MUST have a consistent view of the link protection association for the spans. This can be done using the Link Management Protocol (LMP) [RFC4204], or if LMP is not used, this MUST be configured manually.

2.2. Bi-directional 1+1 Dedicated Protection

Suppose a bi-directional LSP is routed over link i between two nodes A and B. Under bi-directional 1+1 protection, a dedicated link j is pre-assigned to protect the working link i. LSP traffic is permanently duplicated on both links, and under normal conditions, the traffic from link i is received by nodes 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. Note that some form of signaling is required to ensure that both A and B start receiving traffic from the protection link.

The basic steps in 1+1 bi-directional span protection are as follows:

1. If a node (A or B) detects the failure of the working link (or a degradation of signal quality over the working link), it SHOULD begin receiving on the protection link and send a Switchover Request message reliably to the other node (B or A, respectively). This message SHOULD indicate the identity of the failed working link and provide other relevant information.

2. Upon receipt of the Switchover Request message, a node MUST begin receiving from the protection link and send a Switchover Response message to the other node (A or B, respectively). Because both the working/protect spans are exposed to routing and signaling as a single link, the switchover SHOULD be transparent to routing and signaling.

The following functionality is required for 1+1 bi-directional span protection:

- The routing procedures are the same as in 1+1 unidirectional.

- The signaling procedures are the same as in 1+1 unidirectional.

- In addition to the procedures described in 1+1 (unidirectional), a Switchover Request message MUST be used to signal the Switchover Request. This can be done using LMP [RFC4204]. Note that GMPLS-based mechanisms MAY not be necessary when the underlying span (transport) technology provides such a mechanism.

2.3. Dedicated 1:1 Protection with Extra Traffic

Consider two adjacent nodes, A and B. Under 1:1 protection, a dedicated link j between A and B is pre-assigned to protect working link i. Link j may be carrying (pre-emptable) Extra Traffic. A failure affecting link i results in the corresponding LSP(s) being restored to link j. Extra Traffic being routed over link j may need to be pre-empted to accommodate the LSPs that have to be restored.

Once a fault is isolated/localized, the affected LSP(s) must be moved to the protection link. The process of moving an LSP from a failed (working) link to a protection link must be initiated by one of the nodes, A or B. This node is referred to as the "master". The other node is called the "slave". The determination of the master and the slave may be based on configured information or protocol specific requirements.

The basic steps in dedicated 1:1 span protection (ignoring reversion) are as follows:

1. If the master detects/localizes a link failure event, it invokes a process to allocate the protection link to the affected LSP(s).

2. If the slave detects a link failure event, it informs the master of the failure using a failure indication message. The master then invokes the same procedure as (1) to move the LSPs to the protection link. If the protection link is carrying Extra Traffic, the slave stops using the span for the Extra Traffic.

3. Once the span protection procedure is invoked in the master, it requests the slave to switch the affected LSP(s) to the protection link. Prior to this, if the protection link is carrying Extra Traffic, the master stops using the span for this traffic (i.e., the traffic is dropped by the master and not forwarded into or out of the protection link).

4. The slave sends an acknowledgement to the master. Prior to this, the slave stops using the link for Extra Traffic (i.e., the traffic is dropped by the slave and not forwarded into or out of the protection link). It then starts sending the normal traffic on the selected protection link.

5. When the master receives the acknowledgement, it starts sending and receiving the normal traffic over the new link. The switchover of the LSPs is thus completed.

Note: Although this mechanism implies more traffic dropped than necessary, it is preferred over possible misconnections during the recovery process.

From the description above, it is clear that 1:1 span protection may require up to three signaling messages for each failed span: a failure indication message, an LSP Switchover Request message, and an LSP Switchover Response message. Furthermore, it may be possible to switch multiple LSPs from the working span to the protection span simultaneously.

The following functionality is required for dedicated 1:1 span protection:

- Pre-emption MUST be supported to accommodate Extra Traffic.
- Routing: A single TE link encompassing both working and protection links is announced with a Link Protection Type "Dedicated 1:1". If Extra Traffic is supported over the protection link, then the bandwidth parameters for the protection link MUST also be announced. The differentiation between bandwidth for working and protect links is made using priority mechanisms. In other words, the network MUST be configured such that bandwidth at priority X or lower is considered Extra Traffic.

  If there is a failure on the working link, then the normal traffic is switched to the protection link, pre-empting Extra Traffic if necessary. The bandwidth for the protection link MUST be adjusted accordingly.
- Signaling: To establish an LSP on the working link, the Link Protection object/TLV indicating "Dedicated 1:1" SHOULD be included in the signaling request message for that LSP. To establish an LSP on the protection link, the appropriate priority (indicating Extra Traffic) SHOULD be used for that LSP. These objects/TLVs are defined in [RFC3471]. If the Link Protection object/TLV is not used, link selection is a matter of local policy.
- Link management: Both nodes MUST have a consistent view of the link protection association for the spans. This can be done using LMP [RFC4204] or via manual configuration.
- When a link failure is detected at the slave, a failure indication message MUST be sent to the master informing the node of the link failure.

2.4. Shared M:N Protection

Shared M:N protection is described with respect to two neighboring nodes, A and B. The scenario considered is as follows:

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

- When a link in the working set is affected by a failure, the normal traffic is diverted to a link in the protection set, if such a link is available. Note that such a link might be carrying more than one LSP, e.g., an OC-192 link carrying four STS-48 LSPs.

- 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 protection 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 protection links is then subject to policies (e.g., based on relative priority of working traffic). These policies are not specified in this document.

- When normal traffic must be diverted from a failed link in the working set to a protection link, the decision as to which protection link is chosen is always made by one of the nodes, A or B. This node is considered the "master" and it is required to both apply any policies and select specific protection links to divert working traffic. The other node is considered the "slave". The determination of the master and the slave MAY be based on configured information, protocol-specific requirements, or as a result of running a neighbor discovery procedure.

- Failure events are detected by transport layer mechanisms, if available (e.g., SONET Alarm Indication Signal (AIS)/Remote Defect Indication (RDI)). 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 for a

failure to simultaneously affect 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 M:N protection (ignoring reversion) are as follows:

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

2. If the slave detects a failure of a working link, it MUST inform the master of the failure using a failure indication message. The master then invokes the same procedure as above to allocate a protection 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 protection link, it indicates this to the slave and requests the switchover of the traffic (using a "Switchover Request" message). Prior to this, if the protection link is carrying Extra 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 protection link).

4. The slave sends a "Switchover Response" message back to the master. Prior to this, if the selected protection link is carrying traffic that could be pre-empted, 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 protection link). It then starts sending the normal traffic on the selected protection link.

5. When the master receives the Switchover Response, it starts sending and receiving the traffic that was previously carried on the now-failed link over the new link.

Note: Although this mechanism implies more traffic dropped than necessary, it is preferred over possible misconnections during the recovery process.

From the description above, it is clear that M:N span restoration (involving LSP local recovery) MAY require up to three messages for each working link being switched: a failure indication message, a Switchover Request message, and a Switchover Response message.

The following functionality is required for M:N span restoration:

- Pre-emption MUST be supported to accommodate Extra Traffic.
- Routing: A single TE link encompassing both sets of working and protect links should be announced with a Link Protection Type "Shared M:N". If Extra Traffic is supported over a set of the protection links, then the bandwidth parameters for the set of protection links MUST also be announced. The differentiation between bandwidth for working and protect links is made using priority mechanisms.

  If there is a failure on a working link, then the affected LSP(s) MUST be switched to a protection link, pre-empting Extra Traffic if necessary. The bandwidth for the protection link MUST be adjusted accordingly.
- Signaling: To establish an LSP on the working link, the Link Protection object/TLV indicating "Shared M:N" SHOULD be included in the signaling request message for that LSP. To establish an LSP on the protection link, the appropriate priority (indicating Extra Traffic) SHOULD be used. These objects/TLVs are defined in [RFC3471]. If the Link Protection object/TLV is not used, link selection is a matter of local policy.
- For link management, both nodes MUST have a consistent view of the link protection association for the links. This can be done using LMP [RFC4204] or via manual configuration.

2.5. Messages

The following messages are used in local span protection procedures.

These messages SHOULD be delivered reliably. Therefore, the protocol mechanisms used to deliver these messages SHOULD provide sequencing, acknowledgement, and retransmission. The protocol SHOULD also handle situations where the message(s) cannot be delivered.

The messages described in the following subsections are abstract; their format and encoding will be described in separate documents.

2.5.1. Failure Indication Message

This message is sent from the slave to the master to indicate the identities of one or more failed working links. This message MAY not be necessary when the transport plane technology itself provides for such a notification.

The number of links included in the message depends on the number of failures detected within a window of time by the sending node. A node MAY choose to send separate failure indication messages in the interest of completing the recovery for a given link within an implementation-dependent time constraint.

2.5.2. Switchover Request Message

Under bi-directional 1+1 span protection, this message is used to coordinate the selecting function at both nodes. This message originated at the node that detected the failure.

Under dedicated 1:1 and shared M:N span protection, this message is used as an LSP Switchover Request. This message is sent from the master node to the slave node (reliably) to indicate that the LSP(s) on the (failed) working link can be switched to an available protection link. If so, the ID of the protection link, as well as the LSP labels (if necessary), MUST be indicated. These identifiers MUST be consistent with those used in GMPLS signaling.

A working link may carry multiple LSPs. Since the normal traffic carried over the working link is switched to the protection link, it MAY be possible for the LSPs on the working link to be mapped to the protection link without re-signaling each individual LSP. For example, if link bundling [RFC4201] is used where the working and protect links are mapped to component links, and the labels are the same on the working and protection links, it MAY be possible to change the component links without needing to re-signal each individual LSP. Optionally, the labels MAY need to be explicitly coordinated between the two nodes. In this case, the Switchover Request message SHOULD carry the new label mappings.

The master may not be able to find protection links to accommodate all failed working links. Thus, if this message is generated in response to a Failure Indication message from the slave, then the set of failed links in the message MAY be a sub-set of the links received in the Failure Indication message. Depending on time constraints, the master may switch the normal traffic from the set of failed links in smaller batches. Thus, a single failure indication message MAY result in the master sending more than one Switchover Request message to the same slave node.

2.5.3. Switchover Response Message

This message is sent from the slave to the master (reliably) to indicate the completion (or failure) of switchover at the slave. In this message, the slave MAY indicate that it cannot switch over to the corresponding free link for some reason. In this case, the

master and slave notify the user (operator) of the failed switchover. A notification of the failure MAY also be used as a trigger in an end-to-end recovery.

2.6. Preventing Unintended Connections

An unintended connection occurs when traffic from the wrong source is delivered to a receiver. This MUST be prevented during protection switching. This is primarily a concern when the protection link is being used to carry Extra Traffic. In this case, it MUST be ensured that the LSP traffic being switched from the (failed) working link to the protection link is not delivered to the receiver of the pre-empted traffic. Thus, in the message flow described above, the master node MUST disconnect (any) pre-empted traffic on the selected protection link before sending the Switchover Request. The slave node MUST also disconnect pre-empted traffic before sending the Switchover Response. In addition, the master node SHOULD start receiving traffic for the protected LSP from the protection link. Finally, the master node SHOULD start sending protected traffic on the protection link upon receipt of the Switchover Response.

3. End-to-End (Path) Protection and Restoration

End-to-end path protection and restoration refer to the recovery of an entire LSP from the initiator to the terminator. Suppose the primary path of an LSP is routed from the initiator (Node A) to the terminator (Node B) through a set of intermediate nodes.

The following subsections describe three previously proposed end-to-end protection schemes and the functional steps needed to implement them.

3.1. Unidirectional 1+1 Protection

A dedicated, resource-disjoint alternate path is pre-established to protect the LSP. Traffic is simultaneously sent on both paths and received from one of the functional paths by the end nodes A and B.

There is no explicit signaling involved with this mode of protection.

3.2. Bi-directional 1+1 Protection

A dedicated, resource-disjoint alternate path is pre-established to protect the LSP. Traffic is simultaneously sent on both paths; under normal conditions, the traffic from the working path is received by nodes A and B (in the appropriate directions). A failure affecting the working path results in both A and B switching to the traffic on the protection path in the respective directions.

Note that this requires coordination between the end nodes to switch to the protection path.

The basic steps in bi-directional 1+1 path protection are as follows:

- Failure detection: There are two possibilities for this.

  1. A node in the working path detects a failure event. Such a node MUST send a Failure Indication message toward the upstream or/and downstream end node of the LSP (node A or B). This message MAY be forwarded along the working path or routed over a different path if the network has general routing intelligence.

     Mechanisms provided by the data transport plane MAY also be used for this, if available.

  2. The end nodes (A or B) detect the failure themselves (e.g., loss of signal).

- Switchover: The action taken when an end node detects a failure in the working path is as follows: Start receiving from the protection path; at the same time, send a Switchover Request message to the other end node to enable switching at the other end.

  The action taken when an end node receives a Switchover Request message is as follows:

  - Start receiving from the protection path; at the same time, send a Switchover Response message to the other end node.

GMPLS signaling mechanisms MAY be used to (reliably) signal the Failure Indication message, as well as the Switchover Request and Response message. These messages MAY be forwarded along the protection path if no other routing intelligence is available in the network.

3.2.1. Identifiers

LSP Identifier: A unique identifier for each LSP. The LSP identifier is within the scope of the Source ID and Destination ID.

Source ID: ID of the source (e.g., IP address).

Destination ID: ID of the destination (e.g., IP address).

3.2.2. Nodal Information

Each node that is on the working or protection path of an LSP MUST have knowledge of the LSP identifier. If the network does not provide routing intelligence, nodal information MAY also include previous and next nodes in the LSP so that restoration-related messages can be forwarded properly. When the network provides general routing intelligence, messages MAY be forwarded along paths other than that of the LSP.

At the end-point nodes, the working and protection paths MUST be associated. The association of these paths MAY be either provisioned using signaling or MAY be configured when LSP provisioning does not involve signaling (e.g., provisioning through a management system). The related association information MUST remain until the LSP is explicitly de-provisioned.

3.2.3. End-to-End Failure Indication Message

This message is sent (reliably) by an intermediate node toward the source of an LSP. For instance, such a node might have attempted local span protection and failed. This message MAY not be necessary if the data transport layer provides mechanisms for the notification of LSP failure by the endpoints (i.e., if LSP endpoints are co-located with a corresponding data (transport) maintenance/recovery domain).

Consider a node that detects a link failure. The node MUST determine the identities of all LSPs that are affected by the failure of the link and send an End-to-End Failure Indication message to the source of each LSP. For scalability reasons, Failure Indication messages MAY contain the identity and the status of multiple LSPs rather than a single one. Each intermediate node receiving such a message MUST forward the message to the appropriate next node such that the message would ultimately reach the LSP source. However, there is no requirement that this message flows toward the source along the same path as the failed LSP. Furthermore, if an intermediate node is itself generating a Failure Indication message, there SHOULD be a mechanism to suppress all but one source of Failure Indication messages. Finally, the Failure Indication message MUST be sent reliably from the node detecting the failure to the LSP source. Reliability MAY be achieved, for example, by retransmitting the message until an acknowledgement is received. However, retransmission of Failure Indication messages SHOULD not cause further message drops. This MAY be achieved through the appropriate configuration and use of congestion and flow control mechanisms.

3.2.4. End-to-End Failure Acknowledgement Message

This message is sent by the source node to acknowledge the receipt of an End-to-End Failure Indication message. This message is sent to the originator of the Failure Indication message. The Acknowledge message SHOULD be sent for each Failure Indication Message received. Each intermediate node receiving the Failure Acknowledgement message MUST forward it toward the destination of the message. However, there is no requirement that this message flows toward the destination along the same path as the failed LSP.

This message MAY not be required if other means of ensuring reliable message delivery are used.

3.2.5. End-to-End Switchover Request Message

This message is generated by the source node receiving an indication of failure in an LSP. It is sent to the LSP destination, and it carries the identifier of the LSP being restored. The End-to-End Switchover Request message MUST be sent reliably from the source to the destination of the LSP.

3.2.6. End-to-End Switchover Response Message

This message is sent by the destination node receiving an End-to-End Switchover Request message toward the source of the LSP. This message SHOULD identify the LSP being switched over. This message MUST be transmitted in response to each End-to-End Switchover Request message received and MAY indicate either a positive or negative outcome.

3.3. Shared Mesh Restoration

Shared mesh restoration refers to schemes under which protection paths for multiple LSPs share common link and node resources. Under these schemes, the protection capacity is pre-reserved, i.e., link capacity is allocated to protect one or more LSPs, but explicit action is required to instantiate a specific protection LSP. This requires restoration signaling along the protection path. Typically, the protection capacity is shared only amongst LSPs whose working paths are physically diverse. This criterion can be enforced when provisioning the protection path. Specifically, provisioning-related signaling messages may carry information about the working path to nodes along the protection path. This can be used as call admission control to accept/reject connections along the protection path based on the identification of the resources used for the primary path.

Thus, shared mesh restoration is designed to protect an LSP after a single failure event, i.e., a failure that affects the working path of at most one LSP sharing the protection capacity. It is possible that a protection path may not be successfully activated when multiple, concurrent failure events occur. In this case, shared mesh restoration capacity may be claimed for more than one failed LSP and the protection path can be activated only for one of them (at most).

For implementing shared mesh restoration, the identifier and nodal information related to signaling along the control path are as defined for 1+1 protection in Sections 3.2.1 and 3.2.2. In addition, each node MUST also keep (local) information needed to establish the data plane of the protection path. This information MUST indicate the local resources to be allocated, the fabric cross-connect to be established to activate the path, etc. The precise nature of this information would depend on the type of node and LSP (the GMPLS signaling document describes different type of switches [RFC3471]). It would also depend on whether the information is fine or coarse-grained. For example, fine-grained information would indicate pre-selection of all details pertaining to protection path activation, such as outgoing link, labels, etc. Coarse-grained information, on the other hand, would allow some details to be determined during protection path activation. For example, protection resources may be pre-selected at the level of a TE link, while the selection of the specific component link and label occurs during protection path activation.

While the coarser specification allows some flexibility in the selection of the precise resource to activate, it also adds complexity in decision making and signaling during the time-critical restoration phase. Furthermore, the procedures for the assignment of bandwidth to protection paths MUST take into account the total resources in a TE link so that single-failure survivability requirements are satisfied.

3.3.1. End-to-End Failure Indication and Acknowledgement Message

The End-to-End failure indication and acknowledgement procedures and messages are as defined in Sections 3.2.3 and 3.2.4.

3.3.2. End-to-End Switchover Request Message

This message is generated by the source node receiving an indication of failure in an LSP. It is sent to the LSP destination along the protection path, and it identifies the LSP being restored. If any intermediate node is unable to establish cross-connects for the protection path, then it is desirable that no other node in the path

establishes cross-connects for the path. This would allow shared mesh restoration paths to be efficiently utilized.

The End-to-End Switchover message MUST be sent reliably from the source to the destination of the LSP along the protection path.

3.3.3. End-to-End Switchover Response Message

This message is sent by the destination node receiving an End-to-End Switchover Request message toward the source of the LSP, along the protection path. This message SHOULD identify the LSP that is being switched over. Prior to activating the secondary bandwidth at each hop along the path, Extra Traffic (if used) MUST be dropped and not forwarded.

This message MUST be transmitted in response to each End-to-End Switchover Request message received.

4. Reversion and Other Administrative Procedures

Reversion refers to the process of moving an LSP back to the original working path after a failure is cleared and the path is repaired. Reversion applies both to local span and end-to-end path-protected LSPs. Reversion is desired for the following reasons. First, the protection path may not be optimal in comparison to the working path from a routing and resource consumption point of view. Second, moving an LSP to its working path allows the protection resources to be used to protect other LSPs. Reversion has the disadvantage of causing a second service disruption. Use of reversion is at the option of the operator. Reversion implies that a working path remains allocated to the LSP that was originally routed over it, even after a failure. It is important to have mechanisms that allow reversion to be performed with minimal service disruption to the customer. This can be achieved using a "bridge-and-switch" approach (often referred to as make-before-break).

The basic steps involved in bridge-and-switch are as follows:

1. The source node commences the process by "bridging" the normal traffic onto both the working and the protection paths (or links in the case of span protection).

2. Once the bridging process is complete, the source node sends a Bridge and Switch Request message to the destination, identifying the LSP and other information necessary to perform reversion. Upon receipt of this message, the destination

selects the traffic from the working path. At the same time, it bridges the transmitted traffic onto both the working and protection paths.

3. The destination then sends a Bridge and Switch Response message to the source confirming the completion of the operation.

4. When the source receives this message, it switches to receive from the working path, and stops transmitting traffic on the protection path. The source then sends a Bridge and Switch Completed message to the destination confirming that the LSP has been reverted.

5. Upon receipt of this message, the destination stops transmitting along the protection path and de-activates the LSP along this path. The de-activation procedure should remove the crossed connections along the protection path (and frees the resources to be used for restoring other failures).

Administrative procedures other than reversion include the ability to force a switchover (from working to protection or vice versa) and locking out switchover, i.e., preventing an LSP from moving from working to protection administratively. These administrative conditions have to be supported by signaling.

5. Discussion

5.1. LSP Priorities During Protection

Under span protection, a failure event could affect more than one working link and there could be fewer protection links than the number of failed working links. Furthermore, a working link may contain multiple LSPs of varying priority. Under this scenario, a decision must be made as to which working links (and therefore LSPs) should be protected. This decision MAY be based on LSP priorities.

In general, a node 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, LSPs on a working link MAY be switched over to a given protection link, but another failure (of a working link carrying higher priority LSPs) may be detected soon afterward. In this case, the new LSPs may bump the ones previously switched over the protection link.

In the case of end-to-end shared mesh restoration, priorities MAY be implemented for allocating shared link resources under multiple failure scenarios. As described in Section 3.3, more than one LSP

can claim shared resources under multiple failure scenarios. If such resources are first allocated to a lower-priority LSP, they MAY have to be reclaimed and allocated to a higher-priority LSP.

6. Security Considerations

There are a number of security threats that MAY be experienced due to the exchange of messages and information, as detailed in this document. Some examples include interception, spoofing, modification, and replay of control messages. Therefore, the following security requirements are applicable to the mechanisms of this document.

- Signaling MUST be able to provide authentication, integrity, and protection against replay attacks.
- Privacy and confidentiality are not required. Only authentication is required to ensure that the signaling messages are originating from the right place and have not been modified in transit.
- Protection of the identity of the data plane end-points (in Failure Indication messages) is not required

The consequences of poorly secured protection may increase the risk of triggering recovery actions under false Failure Indication messages, including LSP identifiers that are not under failure. Such information could subsequently trigger the initiation of "false" recovery actions while there are no reasons to do so. Additionally, if the identification of the LSP is tampered with from a Failure Indication message, recovery actions will involve nodes for which the LSPs do not indicate any failure condition or for which no Failure Indication message has been received. The consequences of such actions is unpredictable and MAY lead to de-synchronisation between the control and the data plane, as well as increase the risk of misconnections. Moreover, the consequences of poorly applied protection may increase the risk of misconnection. In particular, when Extra Traffic is involved, it is easily possible to deliver the wrong traffic to the wrong destination. Similarly, an intrusion that sets up what appears to be a valid protection LSP and then causes a fault may be able to divert traffic.

Moreover, tampering with a routing information exchange may also have an effect on traffic engineering. Therefore, any mechanisms used for securing and authenticating the transmission of routing information SHOULD be applied in the present context.

7. Contributors

This document was the product of many individuals working together in the CCAMP WG Protection and Restoration design team. The following are the authors that contributed to this document:

Deborah Brungard (AT&T)
200 S. Laurel Ave.
Middletown, NJ 07748, USA

EMail: dbrungard@att.com

Sudheer Dharanikota

EMail: sudheer@ieee.org

Jonathan P. Lang (Sonos)
223 East De La Guerra Street
Santa Barbara, CA 93101, USA

EMail: jplang@ieee.org

Guangzhi Li (AT&T)
180 Park Avenue,
Florham Park, NJ 07932, USA

EMail: gli@research.att.com

Eric Mannie

EMail: eric_mannie@hotmail.com

Dimitri Papadimitriou (Alcatel)
Francis Wellesplein, 1
B-2018 Antwerpen, Belgium

EMail: dimitri.papadimitriou@alcatel.be

Bala Rajagopalan
Microsoft India Development Center
Hyderabad, India

EMail: balar@microsoft.com

Yakov Rekhter (Juniper)
1194 N. Mathilda Avenue
Sunnyvale, CA 94089, USA

EMail: yakov@juniper.net

8. References

8.1. Normative References

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

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

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

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

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

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

8.2. Informative References

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

[RFC4427] Mannie, E., Ed. and D. Papadimitriou, Ed., "Recovery (Protection and Restoration) Terminology for Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4427, March 2006.

Editors' Addresses

Jonathan P. Lang
Sonos, Inc.
223 East De La Guerra Street
Santa Barbara, CA 93101

EMail: jplang@ieee.org

Bala Rajagopalan
Microsoft India Development Center
Hyderabad, India

Ph: +91-40-5502-7423
EMail: balar@microsoft.com

Dimitri Papadimitriou
Alcatel
Francis Wellesplein, 1
B-2018 Antwerpen, Belgium

Phone: +32 3 240-8491
EMail: dimitri.papadimitriou@alcatel.be

Full Copyright Statement

Intellectual Property

Acknowledgement

Funding for the RFC Editor function is provided by the IETF Administrative Support Activity (IASA).