

## **Analysis of Generalized Multi-Protocol Label Switching (GMPLS)-based Recovery Mechanisms (including Protection and Restoration)**

### Status of This Memo

This memo provides information for the Internet community. It does not specify an Internet standard of any kind. Distribution of this memo is unlimited.

### Copyright Notice

Copyright (C) The Internet Society (2006).

### Abstract

This document provides an analysis grid to evaluate, compare, and contrast the Generalized Multi-Protocol Label Switching (GMPLS) protocol suite capabilities with the recovery mechanisms currently proposed at the IETF CCAMP Working Group. A detailed analysis of each of the recovery phases is provided using the terminology defined in [RFC 4427](#). This document focuses on transport plane survivability and recovery issues and not on control plane resilience and related aspects.

### Table of Contents

<a href="#">1. Introduction</a>	<a href="#">3</a>
<a href="#">2. Contributors</a>	<a href="#">4</a>
<a href="#">3. Conventions Used in this Document</a>	<a href="#">5</a>
<a href="#">4. Fault Management</a>	<a href="#">5</a>
<a href="#">4.1. Failure Detection</a>	<a href="#">5</a>
<a href="#">4.2. Failure Localization and Isolation</a>	<a href="#">8</a>
<a href="#">4.3. Failure Notification</a>	<a href="#">9</a>
<a href="#">4.4. Failure Correlation</a>	<a href="#">11</a>
<a href="#">5. Recovery Mechanisms</a>	<a href="#">11</a>
<a href="#">5.1. Transport vs. Control Plane Responsibilities</a>	<a href="#">11</a>
5.2. Technology-Independent and Technology-Dependent Mechanisms	<a href="#">12</a>
<a href="#">5.2.1. OTN Recovery</a>	<a href="#">12</a>
<a href="#">5.2.2. Pre-OTN Recovery</a>	<a href="#">13</a>
<a href="#">5.2.3. SONET/SDH Recovery</a>	<a href="#">13</a>

5.3. Specific Aspects of Control Plane-Based Recovery Mechanisms .....	14
5.3.1. In-Band vs. Out-Of-Band Signaling .....	14
5.3.2. Uni- vs. Bi-Directional Failures .....	15
5.3.3. Partial vs. Full Span Recovery .....	17
5.3.4. Difference between LSP, LSP Segment and Span Recovery .....	18
5.4. Difference between Recovery Type and Scheme .....	19
5.5. LSP Recovery Mechanisms .....	21
5.5.1. Classification .....	21
5.5.2. LSP Restoration .....	23
5.5.3. Pre-Planned LSP Restoration .....	24
5.5.4. LSP Segment Restoration .....	25
6. Reversion .....	26
6.1. Wait-To-Restore (WTR) .....	26
6.2. Revertive Mode Operation .....	26
6.3. Orphans .....	27
7. Hierarchies .....	27
7.1. Horizontal Hierarchy (Partitioning) .....	28
7.2. Vertical Hierarchy (Layers) .....	28
7.2.1. Recovery Granularity .....	30
7.3. Escalation Strategies .....	30
7.4. Disjointness .....	31
7.4.1. SRLG Disjointness .....	32
8. Recovery Mechanisms Analysis .....	33
8.1. Fast Convergence (Detection/Correlation and Hold-off Time) .....	34
8.2. Efficiency (Recovery Switching Time) .....	34
8.3. Robustness .....	35
8.4. Resource Optimization .....	36
8.4.1. Recovery Resource Sharing .....	37
8.4.2. Recovery Resource Sharing and SRLG Recovery .....	39
8.4.3. Recovery Resource Sharing, SRLG Disjointness and Admission Control .....	40
9. Summary and Conclusions .....	42
10. Security Considerations .....	43
11. Acknowledgements .....	43
12. References .....	44
12.1. Normative References .....	44
12.2. Informative References .....	44



## 1. Introduction

This document provides an analysis grid to evaluate, compare, and contrast the Generalized MPLS (GMPLS) protocol suite capabilities with the recovery mechanisms proposed at the IETF CCAMP Working Group. The focus is on transport plane survivability and recovery issues and not on control-plane-resilience-related aspects. Although the recovery mechanisms described in this document impose different requirements on GMPLS-based recovery protocols, the protocols' specifications will not be covered in this document. Though the concepts discussed are technology independent, this document implicitly focuses on SONET [[T1.105](#)]/SDH [[G.707](#)], Optical Transport Networks (OTN) [[G.709](#)], and pre-OTN technologies, except when specific details need to be considered (for instance, in the case of failure detection).

A detailed analysis is provided for each of the recovery phases as identified in [[RFC4427](#)]. These phases define the sequence of generic operations that need to be performed when a LSP/Span failure (or any other event generating such failures) occurs:

- Phase 1: Failure Detection
- Phase 2: Failure Localization (and Isolation)
- Phase 3: Failure Notification
- Phase 4: Recovery (Protection or Restoration)
- Phase 5: Reversion (Normalization)

Together, failure detection, localization, and notification phases are referred to as "fault management". Within a recovery domain, the entities involved during the recovery operations are defined in [[RFC4427](#)]; these entities include ingress, egress, and intermediate nodes. The term "recovery mechanism" is used to cover both protection and restoration mechanisms. Specific terms such as "protection" and "restoration" are used only when differentiation is required. Likewise the term "failure" is used to represent both signal failure and signal degradation.

In addition, when analyzing the different hierarchical recovery mechanisms including disjointness-related issues, a clear distinction is made between partitioning (horizontal hierarchy) and layering (vertical hierarchy). In order to assess the current GMPLS protocol capabilities and the potential need for further extensions, the dimensions for analyzing each of the recovery mechanisms detailed in this document are introduced. This document concludes by detailing the applicability of the current GMPLS protocol building blocks for recovery purposes.



## 2. Contributors

This document is the result of the CCAMP Working Group Protection and Restoration design team joint effort. Besides the editors, the following are the authors that contributed to the present memo:

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

EMail: [dbrungard@att.com](mailto:dbrungard@att.com)

Sudheer Dharanikota

EMail: [sudheer@ieee.org](mailto:sudheer@ieee.org)

Jonathan P. Lang (Sonos)  
506 Chapala Street  
Santa Barbara, CA 93101, USA

EMail: [jplang@ieee.org](mailto:jplang@ieee.org)

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

EMail: [gli@research.att.com](mailto:gli@research.att.com)

Eric Mannie  
Perceval  
Rue Tenbosch, 9  
1000 Brussels  
Belgium

Phone: +32-2-6409194  
EMail: [eric.mannie@perceval.net](mailto:eric.mannie@perceval.net)

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

EMail: [dimitri.papadimitriou@alcatel.be](mailto:dimitri.papadimitriou@alcatel.be)



Bala Rajagopalan  
Microsoft India Development Center  
Hyderabad, India

EEmail: balar@microsoft.com

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

EEmail: yakov@juniper.net

### **3. 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 [\[RFC2119\]](#).

Any other recovery-related terminology used in this document conforms to that defined in [\[RFC4427\]](#). The reader is also assumed to be familiar with the terminology developed in [\[RFC3945\]](#), [\[RFC3471\]](#), [\[RFC3473\]](#), [\[RFC4202\]](#), and [\[RFC4204\]](#).

### **4. Fault Management**

#### **4.1. Failure Detection**

Transport failure detection is the only phase that cannot be achieved by the control plane alone because the latter needs a hook to the transport plane in order to collect the related information. It has to be emphasized that even if failure events themselves are detected by the transport plane, the latter, upon a failure condition, must trigger the control plane for subsequent actions through the use of GMPLS signaling capabilities (see [\[RFC3471\]](#) and [\[RFC3473\]](#)) or Link Management Protocol capabilities (see [\[RFC4204\]](#), [Section 6](#)).

Therefore, by definition, transport failure detection is transport technology dependent (and so exceptionally, we keep here the "transport plane" terminology). In transport fault management, distinction is made between a defect and a failure. Here, the discussion addresses failure detection (persistent fault cause). In the technology-dependent descriptions, a more precise specification will be provided.

As an example, SONET/SDH (see [\[G.707\]](#), [\[G.783\]](#), and [\[G.806\]](#)) provides supervision capabilities covering:





- Continuity: SONET/SDH monitors the integrity of the continuity of a trail (i.e., section or path). This operation is performed by monitoring the presence/absence of the signal. Examples are Loss of Signal (LOS) detection for the physical layer, Unequipped (UNEQ) Signal detection for the path layer, Server Signal Fail Detection (e.g., AIS) at the client layer.
- Connectivity: SONET/SDH monitors the integrity of the routing of the signal between end-points. Connectivity monitoring is needed if the layer provides flexible connectivity, either automatically (e.g., cross-connects) or manually (e.g., fiber distribution frame). An example is the Trail (i.e., section or path) Trace Identifier used at the different layers and the corresponding Trail Trace Identifier Mismatch detection.
- Alignment: SONET/SDH checks that the client and server layer frame start can be correctly recovered from the detection of loss of alignment. The specific processes depend on the signal/frame structure and may include: (multi-)frame alignment, pointer processing, and alignment of several independent frames to a common frame start in case of inverse multiplexing. Loss of alignment is a generic term. Examples are loss of frame, loss of multi-frame, or loss of pointer.
- Payload type: SONET/SDH checks that compatible adaptation functions are used at the source and the destination. Normally, this is done by adding a payload type identifier (referred to as the "signal label") at the source adaptation function and comparing it with the expected identifier at the destination. For instance, the payload type identifier is compared with the corresponding mismatch detection.
- Signal Quality: SONET/SDH monitors the performance of a signal. For instance, if the performance falls below a certain threshold, a defect -- excessive errors (EXC) or degraded signal (DEG) -- is detected.

The most important point is that the supervision processes and the corresponding failure detection (used to initiate the recovery phase(s)) result in either:

- Signal Degrade (SD): A signal indicating that the associated data has degraded in the sense that a degraded defect condition is active (for instance, a dDEG declared when the Bit Error Rate exceeds a preset threshold). Or



- Signal Fail (SF): A signal indicating that the associated data has failed in the sense that a signal interrupting near-end defect condition is active (as opposed to the degraded defect).

In Optical Transport Networks (OTN), equivalent supervision capabilities are provided at the optical/digital section layers (i.e., Optical Transmission Section (OTS), Optical Multiplex Section (OMS) and Optical channel Transport Unit (OTU)) and at the optical/digital path layers (i.e., Optical Channel (OCh) and Optical channel Data Unit (ODU)). Interested readers are referred to the ITU-T Recommendations [[G.798](#)] and [[G.709](#)] for more details.

The above are examples that illustrate cases where the failure detection and reporting entities (see [[RFC4427](#)]) are co-located. The following example illustrates the scenario where the failure detecting and reporting entities (see [[RFC4427](#)]) are not co-located.

In pre-OTN networks, a failure may be masked by intermediate O-E-O based Optical Line System (OLS), preventing a Photonic Cross-Connect (PXC) from detecting upstream failures. In such cases, failure detection may be assisted by an out-of-band communication channel, and failure condition may be reported to the PXC control plane. This can be provided by using [[RFC4209](#)] extensions that deliver IP message-based communication between the PXC and the OLS control plane. Also, since PXC's are independent of the framing format, failure conditions can only be triggered either by detecting the absence of the optical signal or by measuring its quality. These mechanisms are generally less reliable than electrical (digital) ones. Both types of detection mechanisms are outside the scope of this document. If the intermediate OLS supports electrical (digital) mechanisms, using the LMP communication channel, these failure conditions are reported to

the PXC and subsequent recovery actions are performed as described in [Section 5](#). As such, from the control plane viewpoint, this mechanism turns the OLS-PXC-composed system into a single logical entity, thus having the same failure management mechanisms as any other O-E-O capable device.

More generally, the following are typical failure conditions in SONET/SDH and pre-OTN networks:

- Loss of Light (LOL)/Loss of Signal (LOS): Signal Failure (SF) condition where the optical signal is not detected any longer on the receiver of a given interface.
- Signal Degrade (SD): detection of the signal degradation over a specific period of time.



- For SONET/SDH payloads, all of the above-mentioned supervision capabilities can be used, resulting in SD or SF conditions.

In summary, the following cases apply when considering the communication between the detecting and reporting entities:

- Co-located detecting and reporting entities: both the detecting and reporting entities are on the same node (e.g., SONET/SDH equipment, Opaque cross-connects, and, with some limitations, Transparent cross-connects, etc.)
- Non-co-located detecting and reporting entities:
  - o with in-band communication between entities: entities are physically separated, but the transport plane provides in-band communication between them (e.g., Server Signal Failures such as Alarm Indication Signal (AIS), etc.)
  - o with out-of-band communication between entities: entities are physically separated, but an out-of-band communication channel is provided between them (e.g., using [RFC4204]).

#### **4.2. Failure Localization and Isolation**

Failure localization provides information to the deciding entity about the location (and so the identity) of the transport plane entity that detects the LSP(s)/span(s) failure. The deciding entity can then make an accurate decision to achieve finer grained recovery switching action(s). Note that this information can also be included as part of the failure notification (see [Section 4.3](#)).

In some cases, this accurate failure localization information may be less urgent to determine if it requires performing more time-consuming failure isolation (see also [Section 4.4](#)). This is particularly the case when edge-to-edge LSP recovery is performed based on a simple failure notification (including the identification of the working LSPs under failure condition). Note that "edge" refers to a sub-network end-node, for instance. In this case, a more accurate localization and isolation can be performed after recovery of these LSPs.

Failure localization should be triggered immediately after the fault detection phase. This operation can be performed at the transport plane and/or (if the operation is unavailable via the transport plane) the control plane level where dedicated signaling messages can be used. When performed at the control plane level, a protocol such as LMP (see [\[RFC4204\], Section 6](#)) can be used for failure localization purposes.



### 4.3. Failure Notification

Failure notification is used 1) to inform intermediate nodes that an LSP/span failure has occurred and has been detected and 2) to inform the deciding entities (which can correspond to any intermediate or end-point of the failed LSP/span) that the corresponding service is not available. In general, these deciding entities will be the ones making the appropriate recovery decision. When co-located with the recovering entity, these entities will also perform the corresponding recovery action(s).

Failure notification can be provided either by the transport or by the control plane. As an example, let us first briefly describe the failure notification mechanism defined at the SONET/SDH transport plane level (also referred to as maintenance signal supervision):

- AIS (Alarm Indication Signal) occurs as a result of a failure condition such as Loss of Signal and is used to notify downstream nodes (of the appropriate layer processing) that a failure has occurred. AIS performs two functions: 1) inform the intermediate nodes (with the appropriate layer monitoring capability) that a failure has been detected and 2) notify the connection end-point that the service is no longer available.

For a distributed control plane supporting one (or more) failure notification mechanism(s), regardless of the mechanism's actual implementation, the same capabilities are needed with more (or less) information provided about the LSPs/spans under failure condition, their detailed statuses, etc.

The most important difference between these mechanisms is related to the fact that transport plane notifications (as defined today) would directly initiate either a certain type of protection switching (such as those described in [RFC4427]) via the transport plane or restoration actions via the management plane.

On the other hand, using a failure notification mechanism through the control plane would provide the possibility of triggering either a protection or a restoration action via the control plane. This has the advantage that a control-plane-recovery-responsible entity does not necessarily have to be co-located with a transport maintenance/recovery domain. A control plane recovery domain can be defined at entities not supporting a transport plane recovery.

Moreover, as specified in [RFC3473], notification message exchanges through a GMPLS control plane may not follow the same path as the LSP/spans for which these messages carry the status. In turn, this ensures a fast, reliable (through acknowledgement and the use of





either a dedicated control plane network or disjoint control channels), and efficient (through the aggregation of several LSP/span statuses within the same message) failure notification mechanism.

The other important properties to be met by the failure notification mechanism are mainly the following:

- Notification messages must provide enough information such that the most efficient subsequent recovery action will be taken at the recovering entities (in most of the recovery types and schemes this action is even deterministic). Remember here that these entities can be either intermediate or end-points through which normal traffic flows. Based on local policy, intermediate nodes may not use this information for subsequent recovery actions (see for instance the APS protocol phases as described in [\[RFC4427\]](#)). In addition, since fast notification is a mechanism running in collaboration with the existing GMPLS signaling (see [\[RFC3473\]](#)) that also allows intermediate nodes to stay informed about the status of the working LSP/spans under failure condition.

The trade-off here arises when defining what information the LSP/span end-points (more precisely, the deciding entities) need in order for the recovering entity to take the best recovery action: If not enough information is provided, the decision cannot be optimal (note that in this eventuality, the important issue is to quantify the level of sub-optimality). If too much information is provided, the control plane may be overloaded with unnecessary information and the aggregation/correlation of this notification information will be more complex and time-consuming to achieve. Note that a more detailed quantification of the amount of information to be exchanged and processed is strongly dependent on the failure notification protocol.

- If the failure localization and isolation are not performed by one of the LSP/span end-points or some intermediate points, the points should receive enough information from the notification message in order to locate the failure. Otherwise, they would need to (re-) initiate a failure localization and isolation action.
- Avoiding so-called notification storms implies that 1) the failure detection output is correlated (i.e., alarm correlation) and aggregated at the node detecting the failure(s), 2) the failure notifications are directed to a restricted set of destinations (in general the end-points), and 3) failure notification suppression (i.e., alarm suppression) is provided in order to limit flooding in case of multiple and/or correlated failures detected at several locations in the network.



- Alarm correlation and aggregation (at the failure-detecting node) implies a consistent decision based on the conditions for which a trade-off between fast convergence (at detecting node) and fast notification (implying that correlation and aggregation occurs at receiving end-points) can be found.

#### **4.4. Failure Correlation**

A single failure event (such as a span failure) can cause multiple failure (such as individual LSP failures) conditions to be reported. These can be grouped (i.e., correlated) to reduce the number of failure conditions communicated on the reporting channel, for both in-band and out-of-band failure reporting.

In such a scenario, it can be important to wait for a certain period of time, typically called failure correlation time, and gather all the failures to report them as a group of failures (or simply group failure). For instance, this approach can be provided using LMP-WDM for pre-OTN networks (see [[RFC4209](#)]) or when using Signal Failure/Degrade Group in the SONET/SDH context.

Note that a default average time interval during which failure correlation operation can be performed is difficult to provide since it is strongly dependent on the underlying network topology. Therefore, providing a per-node configurable failure correlation time can be advisable. The detailed selection criteria for this time interval are outside of the scope of this document.

When failure correlation is not provided, multiple failure notification messages may be sent out in response to a single failure (for instance, a fiber cut). Each failure notification message contains a set of information on the failed working resources (for instance, the individual lambda LSP flowing through this fiber). This allows for a more prompt response, but can potentially overload the control plane due to a large amount of failure notifications.

### **5. Recovery Mechanisms**

#### **5.1. Transport vs. Control Plane Responsibilities**

When applicable, recovery resources are provisioned, for both protection and restoration, using GMPLS signaling capabilities. Thus, these are control plane-driven actions (topological and resource-constrained) that are always performed in this context.

The following tables give an overview of the responsibilities taken by the control plane in case of LSP/span recovery:



## 1. LSP/span Protection

- |   |                         |
|---|-------------------------|
| - Phase 1: Failure Detection              | Transport plane         |
| - Phase 2: Failure Localization/Isolation | Transport/Control plane |
| - Phase 3: Failure Notification           | Transport/Control plane |
| - Phase 4: Protection Switching           | Transport/Control plane |
| - Phase 5: Reversion (Normalization)      | Transport/Control plane |

Note: in the context of LSP/span protection, control plane actions can be performed either for operational purposes and/or synchronization purposes (vertical synchronization between transport and control plane) and/or notification purposes (horizontal synchronization between end-nodes at control plane level). This suggests the selection of the responsible plane (in particular for protection switching) during the provisioning phase of the protected/protection LSP.

## 2. LSP/span Restoration

- |   |                         |
|---|-------------------------|
| - Phase 1: Failure Detection              | Transport plane         |
| - Phase 2: Failure Localization/Isolation | Transport/Control plane |
| - Phase 3: Failure Notification           | Control plane           |
| - Phase 4: Recovery Switching             | Control plane           |
| - Phase 5: Reversion (Normalization)      | Control plane           |

Therefore, this document primarily focuses on provisioning of LSP recovery resources, failure notification mechanisms, recovery switching, and reversion operations. Moreover, some additional considerations can be dedicated to the mechanisms associated to the failure localization/isolation phase.

## 5.2. Technology-Independent and Technology-Dependent Mechanisms

The present recovery mechanisms analysis applies to any circuit-oriented data plane technology with discrete bandwidth increments (like SONET/SDH, G.709 OTN, etc.) being controlled by a GMPLS-based distributed control plane.

The following sub-sections are not intended to favor one technology versus another. They list pro and cons for each technology in order to determine the mechanisms that GMPLS-based recovery must deliver to overcome their cons and make use of their pros in their respective applicability context.

### 5.2.1. OTN Recovery

OTN recovery specifics are left for further consideration.



### 5.2.2. Pre-OTN Recovery

Pre-OTN recovery specifics (also referred to as "lambda switching") present mainly the following advantages:

- They benefit from a simpler architecture, making it more suitable for mesh-based recovery types and schemes (on a per-channel basis).
- Failure suppression at intermediate node transponders, e.g., use of squelching, implies that failures (such as LoL) will propagate to edge nodes. Thus, edge nodes will have the possibility to initiate recovery actions driven by upper layers (vs. use of non-standard masking of upstream failures).

The main disadvantage is the lack of interworking due to the large amount of failure management (in particular failure notification protocols) and recovery mechanisms currently available.

Note also, that for all-optical networks, combination of recovery with optical physical impairments is left for a future release of this document because corresponding detection technologies are under specification.

### 5.2.3. SONET/SDH Recovery

Some of the advantages of SONET [[T1.105](#)]/SDH [[G.707](#)], and more generically any Time Division Multiplexing (TDM) transport plane recovery, are that they provide:

- Protection types operating at the data plane level that are standardized (see [[G.841](#)]) and can operate across protected domains and interwork (see [[G.842](#)]).
- Failure detection, notification, and path/section Automatic Protection Switching (APS) mechanisms.
- Greater control over the granularity of the TDM LSPs/links that can be recovered with respect to coarser optical channel (or whole fiber content) recovery switching

Some of the limitations of the SONET/SDH recovery are:

- Limited topological scope: Inherently the use of ring topologies, typically, dedicated Sub-Network Connection Protection (SNCP) or shared protection rings, has reduced flexibility and resource efficiency with respect to the (somewhat more complex) meshed recovery.





- Inefficient use of spare capacity: SONET/SDH protection is largely applied to ring topologies, where spare capacity often remains idle, making the efficiency of bandwidth usage a real issue.
- Support of meshed recovery requires intensive network management development, and the functionality is limited by both the network elements and the capabilities of the element management systems (thus justifying the development of GMPLS-based distributed recovery mechanisms.)

### **5.3. Specific Aspects of Control Plane-Based Recovery Mechanisms**

#### **5.3.1. In-Band vs. Out-Of-Band Signaling**

The nodes communicate through the use of IP terminating control channels defining the control plane (transport) topology. In this context, two classes of transport mechanisms can be considered here: in-fiber or out-of-fiber (through a dedicated physically diverse control network referred to as the Data Communication Network or DCN). The potential impact of the usage of an in-fiber (signaling) transport mechanism is briefly considered here.

In-fiber transport mechanisms can be further subdivided into in-band and out-of-band. As such, the distinction between in-fiber in-band and in-fiber out-of-band signaling reduces to the consideration of a logically- versus physically-embedded control plane topology with respect to the transport plane topology. In the scope of this document, it is assumed that at least one IP control channel between each pair of adjacent nodes is continuously available to enable the exchange of recovery-related information and messages. Thus, in either case (i.e., in-band or out-of-band) at least one logical or physical control channel between each pair of nodes is always expected to be available.

Therefore, the key issue when using in-fiber signaling is whether one can assume independence between the fault-tolerance capabilities of control plane and the failures affecting the transport plane (including the nodes). Note also that existing specifications like the OTN provide a limited form of independence for in-fiber signaling by dedicating a separate optical supervisory channel (OSC, see [\[G.709\]](#) and [\[G.874\]](#)) to transport the overhead and other control traffic. For OTNs, failure of the OSC does not result in failing the optical channels. Similarly, loss of the control channel must not result in failing the data channels (transport plane).



### 5.3.2. Uni- vs. Bi-Directional Failures

The failure detection, correlation, and notification mechanisms (described in [Section 4](#)) can be triggered when either a uni-directional or a bi-directional LSP/Span failure occurs (or a combination of both). As illustrated in Figures 1 and 2, two alternatives can be considered here:

1. Uni-directional failure detection: the failure is detected on the receiver side, i.e., it is detected by only the downstream node to the failure (or by the upstream node depending on the failure propagation direction, respectively).
2. Bi-directional failure detection: the failure is detected on the receiver side of both downstream node AND upstream node to the failure.

Notice that after the failure detection time, if only control-plane-based failure management is provided, the peering node is unaware of the failure detection status of its neighbor.

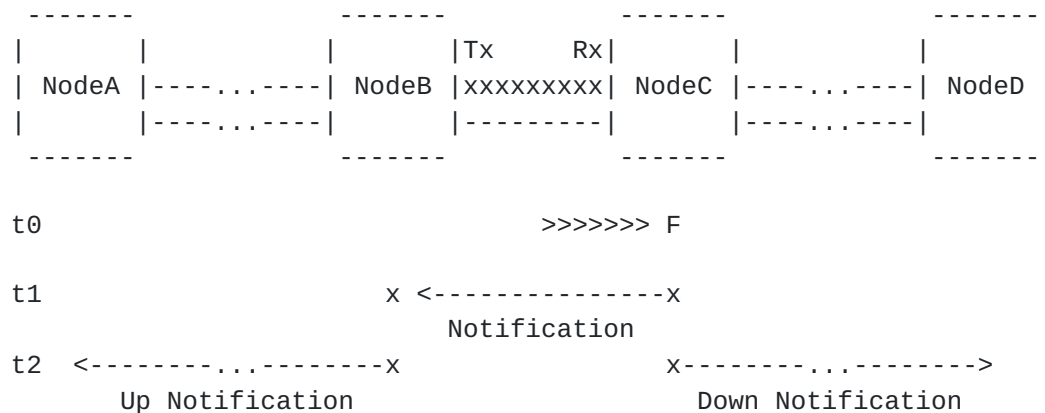


Figure 1: Uni-directional failure detection



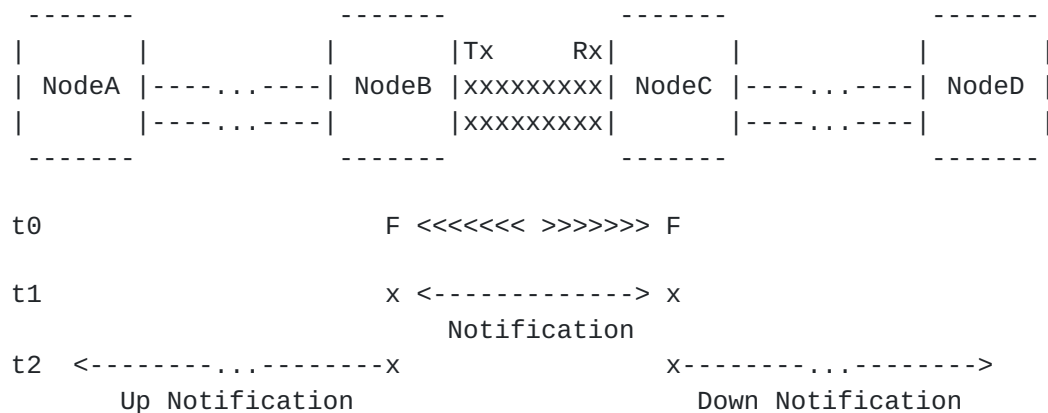


Figure 2: Bi-directional failure detection

After failure detection, the following failure management operations can be subsequently considered:

- Each detecting entity sends a notification message to the corresponding transmitting entity. For instance, in Figure 1, node C sends a notification message to node B. In Figure 2, node C sends a notification message to node B while node B sends a notification message to node C. To ensure reliable failure notification, a dedicated acknowledgement message can be returned back to the sender node.
- Next, within a certain (and pre-determined) time window, nodes impacted by the failure occurrences may perform their correlation. In case of uni-directional failure, node B only receives the notification message from node C, and thus the time for this operation is negligible. In case of bi-directional failure, node B has to correlate the received notification message from node C with the corresponding locally detected information (and node C has to do the same with the message from node B).
- After some (pre-determined) period of time, referred to as the hold-off time, if the local recovery actions (see [Section 5.3.4](#)) were not successful, the following occurs. In case of uni-directional failure and depending on the directionality of the LSP, node B should send an upstream notification message (see [\[RFC3473\]](#)) to the ingress node A. Node C may send a downstream notification message (see [\[RFC3473\]](#)) to the egress node D. However, in that case, only node A would initiate an edge to edge recovery action. Node A is referred to as the "master", and node D is referred to as the "slave", per [\[RFC4427\]](#). Note that the other LSP end-node (node D in this case) may be optionally notified using a downstream notification message (see [\[RFC3473\]](#)).



In case of bi-directional failure, node B should send an upstream notification message (see [RFC3473]) to the ingress node A. Node C may send a downstream notification message (see [RFC3473]) to the egress node D. However, due to the dependence on the LSP directionality, only ingress node A would initiate an edge-to-edge recovery action. Note that the other LSP end-node (node D in this case) should also be notified of this event using a downstream notification message (see [RFC3473]). For instance, if an LSP directed from D to A is under failure condition, only the notification message sent from node C to D would initiate a recovery action. In this case, per [RFC4427], the deciding and recovering node D is referred to as the "master", while node A is referred to as the "slave" (i.e., recovering only entity).

Note: The determination of the master and the slave may be based either on configured information or dedicated protocol capability.

In the above scenarios, the path followed by the upstream and downstream notification messages does not have to be the same as the one followed by the failed LSP (see [RFC3473] for more details on the notification message exchange). The important point concerning this mechanism is that either the detecting/reporting entity (i.e., nodes B and C) is also the deciding/recovery entity or the detecting/reporting entity is simply an intermediate node in the subsequent recovery process. One refers to local recovery in the former case, and to edge-to-edge recovery in the latter one (see also [Section 5.3.4](#)).

### **5.3.3. Partial vs. Full Span Recovery**

When a given span carries more than one LSP or LSP segment, an additional aspect must be considered. In case of span failure, the LSPs it carries can be recovered individually, as a group (aka bulk LSP recovery), or as independent sub-groups. When correlation time windows are used and simultaneous recovery of several LSPs can be performed using a single request, the selection of this mechanism would be triggered independently of the failure notification granularity. Moreover, criteria for forming such sub-groups are outside of the scope of this document.

Additional complexity arises in the case of (sub-)group LSP recovery. Between a given pair of nodes, the LSPs that a given (sub-)group contains may have been created from different source nodes (i.e., initiator) and directed toward different destination nodes. Consequently the failure notification messages following a bi-directional span failure that affects several LSPs (or the whole group of LSPs it carries) are not necessarily directed toward the same initiator nodes. In particular, these messages may be directed





to both the upstream and downstream nodes to the failure. Therefore, such span failure may trigger recovery actions to be performed from both sides (i.e., from both the upstream and the downstream nodes to the failure). In order to facilitate the definition of the corresponding recovery mechanisms (and their sequence), one assumes here as well that, per [RFC4427], the deciding (and recovering) entity (referred to as the "master") is the only initiator of the recovery of the whole LSP (sub-)group.

#### **5.3.4. Difference between LSP, LSP Segment and Span Recovery**

The recovery definitions given in [RFC4427] are quite generic and apply for link (or local span) and LSP recovery. The major difference between LSP, LSP Segment and span recovery is related to the number of intermediate nodes that the signaling messages have to travel. Since nodes are not necessarily adjacent in the case of LSP (or LSP Segment) recovery, signaling message exchanges from the reporting to the deciding/recovery entity may have to cross several intermediate nodes. In particular, this applies to the notification messages due to the number of hops separating the location of a failure occurrence from its destination. This results in an additional propagation and forwarding delay. Note that the former delay may in certain circumstances be non-negligible; e.g., in a copper out-of-band network, the delay is approximately 1 ms per 200km.

Moreover, the recovery mechanisms applicable to end-to-end LSPs and to the segments that may compose an end-to-end LSP (i.e., edge-to-edge recovery) can be exactly the same. However, one expects in the latter case, that the destination of the failure notification message will be the ingress/egress of each of these segments. Therefore, using the mechanisms described in [Section 5.3.2](#), failure notification messages can be exchanged first between terminating points of the LSP segment, and after expiration of the hold-off time, between terminating points of the end-to-end LSP.

Note: Several studies provide quantitative analysis of the relative performance of LSP/span recovery techniques. [WANG] for instance, provides an analysis grid for these techniques showing that dynamic LSP restoration (see [Section 5.5.2](#)) performs well under medium network loads, but suffers performance degradations at higher loads due to greater contention for recovery resources. LSP restoration upon span failure, as defined in [WANG], degrades at higher loads because paths around failed links tend to increase the hop count of the affected LSPs and thus consume additional network resources. Also, performance of LSP restoration can be enhanced by a failed working LSP's source node that initiates a new recovery attempt if an initial attempt fails. A single retry attempt is sufficient to



produce large increases in the restoration success rate and ability to initiate successful LSP restoration attempts, especially at high loads, while not adding significantly to the long-term average recovery time. Allowing additional attempts produces only small additional gains in performance. This suggests using additional (intermediate) crankback signaling when using dynamic LSP restoration (described in [Section 5.5.2](#) - case 2). Details on crankback signaling are outside the scope of this document.

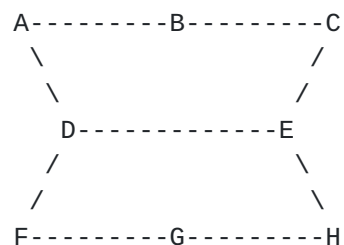
#### 5.4. Difference between Recovery Type and Scheme

[RFC4427] defines the basic LSP/span recovery types. This section describes the recovery schemes that can be built using these recovery types. In brief, a recovery scheme is defined as the combination of several ingress-egress node pairs supporting a given recovery type (from the set of the recovery types they allow). Several examples are provided here to illustrate the difference between recovery types such as 1:1 or M:N, and recovery schemes such as  $(1:1)^n$  or  $(M:N)^n$  (referred to as shared-mesh recovery).

##### 1. $(1:1)^n$ with recovery resource sharing

The exponent,  $n$ , indicates the number of times a 1:1 recovery type is applied between at most  $n$  different ingress-egress node pairs. Here, at most  $n$  pairs of disjoint working and recovery LSPs/spans share a common resource at most  $n$  times. Since the working LSPs/spans are mutually disjoint, simultaneous requests for use of the shared (common) resource will only occur in case of simultaneous failures, which are less likely to happen.

For instance, in the common  $(1:1)^2$  case, if the 2 recovery LSPs in the group overlap the same common resource, then it can handle only single failures; any multiple working LSP failures will cause at least one working LSP to be denied automatic recovery. Consider for instance the following topology with the working LSPs A-B-C and F-G-H and their respective recovery LSPs A-D-E-C and F-D-E-H that share a common D-E link resource.



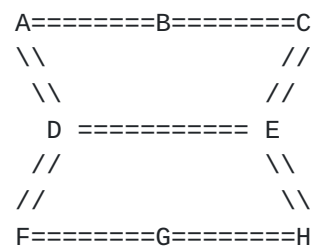


## 2. $(M:N)^n$ with recovery resource sharing

The  $(M:N)^n$  scheme is documented here for the sake of completeness only (i.e., it is not mandated that GMPLS capabilities support this scheme). The exponent,  $n$ , indicates the number of times an  $M:N$  recovery type is applied between at most  $n$  different ingress-egress node pairs. So the interpretation follows from the previous case, except that here disjointness applies to the  $N$  working LSPs/spans and to the  $M$  recovery LSPs/spans while sharing at most  $n$  times  $M$  common resources.

In both schemes, it results in a "group" of  $\sum_{n=1}^N N\{n\}$  working LSPs and a pool of shared recovery resources, not all of which are available to any given working LSP. In such conditions, defining a metric that describes the amount of overlap among the recovery LSPs would give some indication of the group's ability to handle simultaneous failures of multiple LSPs.

For instance, in the simple  $(1:1)^n$  case, if  $n$  recovery LSPs in a  $(1:1)^n$  group overlap, then the group can handle only single failures; any simultaneous failure of multiple working LSPs will cause at least one working LSP to be denied automatic recovery. But if one considers, for instance, a  $(2:2)^2$  group in which there are two pairs of overlapping recovery LSPs, then two LSPs (belonging to the same pair) can be simultaneously recovered. The latter case can be illustrated by the following topology with 2 pairs of working LSPs A-B-C and F-G-H and their respective recovery LSPs A-D-E-C and F-D-E-H that share two common D-E link resources.



Moreover, in all these schemes, (working) path disjointness can be enforced by exchanging information related to working LSPs during the recovery LSP signaling. Specific issues related to the combination of shared (discrete) bandwidth and disjointness for recovery schemes are described in [Section 8.4.2](#).



## **5.5. LSP Recovery Mechanisms**

### **5.5.1. Classification**

The recovery time and ratio of LSPs/spans depend on proper recovery LSP provisioning (meaning pre-provisioning when performed before failure occurrence) and the level of overbooking of recovery resources (i.e., over-provisioning). A proper balance of these two operations will result in the desired LSP/span recovery time and ratio when single or multiple failures occur. Note also that these operations are mostly performed during the network planning phases.

The different options for LSP (pre-)provisioning and overbooking are classified below to structure the analysis of the different recovery mechanisms.

#### **1. Pre-Provisioning**

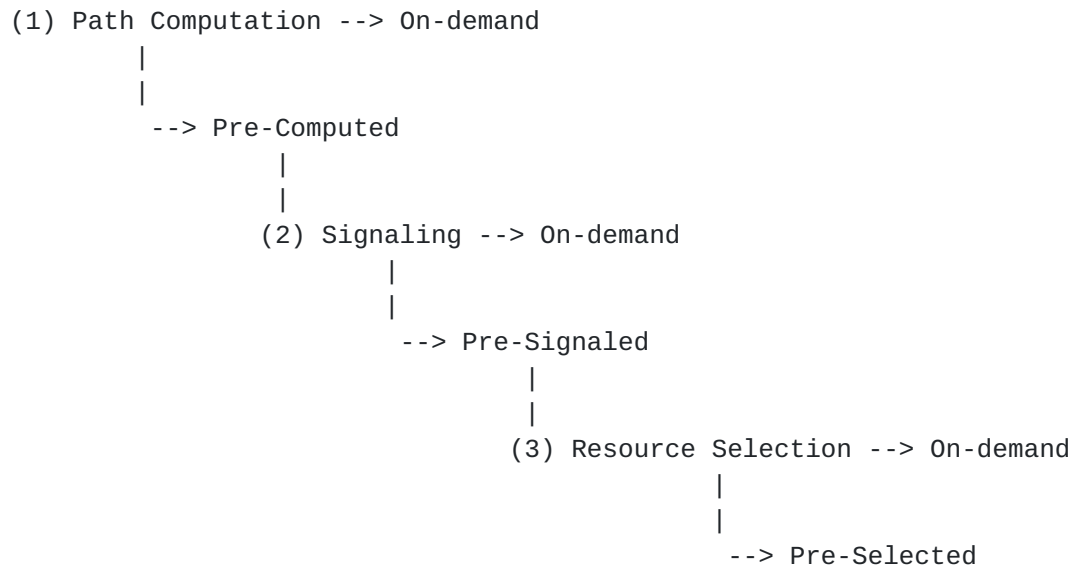
Proper recovery LSP pre-provisioning will help to alleviate the failure of the working LSPs (due to the failure of the resources that carry these LSPs). As an example, one may compute and establish the recovery LSP either end-to-end or segment-per-segment, to protect a working LSP from multiple failure events affecting link(s), node(s) and/or SRLG(s). The recovery LSP pre-provisioning options are classified as follows in the figure below:

- (1) The recovery path can be either pre-computed or computed on-demand.
- (2) When the recovery path is pre-computed, it can be either pre-sigaled (implying recovery resource reservation) or signaled on-demand.
- (3) When the recovery resources are pre-sigaled, they can be either pre-selected or selected on-demand.

Recovery LSP provisioning phases:





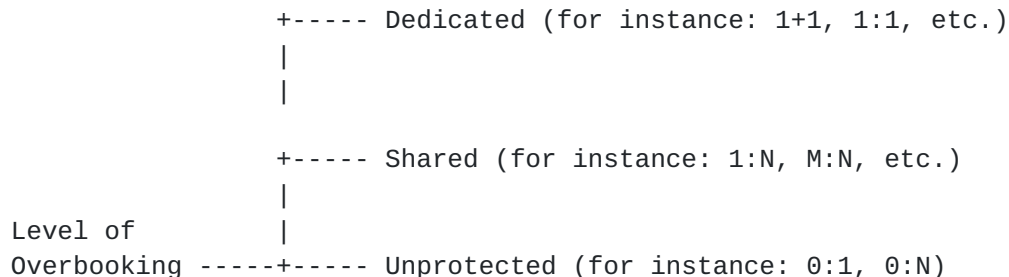


Note that these different options lead to different LSP/span recovery times. The following sections will consider the above-mentioned pre-provisioning options when analyzing the different recovery mechanisms.

## 2. Overbooking

There are many mechanisms available that allow the overbooking of the recovery resources. This overbooking can be done per LSP (as in the example mentioned above), per link (such as span protection), or even per domain. In all these cases, the level of overbooking, as shown in the below figure, can be classified as dedicated (such as 1+1 and 1:1), shared (such as 1:N and M:N), or unprotected (and thus restorable, if enough recovery resources are available).

Overbooking levels:



Also, when using shared recovery, one may support preemptible extra-traffic; the recovery mechanism is then expected to allow preemption of this low priority traffic in case of recovery resource contention during recovery operations. The following sections will consider the



above-mentioned overbooking options when analyzing the different recovery mechanisms.

### 5.5.2. LSP Restoration

The following times are defined to provide a quantitative estimation about the time performance of the different LSP restoration mechanisms (also referred to as LSP re-routing):

- Path Computation Time:  $T_c$
- Path Selection Time:  $T_s$
- End-to-end LSP Resource Reservation Time:  $T_r$  (a delta for resource selection is also considered, the corresponding total time is then referred to as  $T_{rs}$ )
- End-to-end LSP Resource Activation Time:  $T_a$  (a delta for resource selection is also considered, the corresponding total time is then referred to as  $T_{as}$ )

The Path Selection Time ( $T_s$ ) is considered when a pool of recovery LSP paths between a given pair of source/destination end-points is pre-computed, and after a failure occurrence one of these paths is selected for the recovery of the LSP under failure condition.

Note: failure management operations such as failure detection, correlation, and notification are considered (for a given failure event) as equally time-consuming for all the mechanisms described below:

#### 1. With Route Pre-computation (or LSP re-provisioning)

An end-to-end restoration LSP is established after the failure(s) occur(s) based on a pre-computed path. As such, one can define this as an "LSP re-provisioning" mechanism. Here, one or more (disjoint) paths for the restoration LSP are computed (and optionally pre-selected) before a failure occurs.

No reservation or selection of resources is performed along the restoration path before failure occurrence. As a result, there is no guarantee that a restoration LSP is available when a failure occurs.

The expected total restoration time  $T$  is thus equal to  $T_s + T_{rs}$  or to  $T_{rs}$  when a dedicated computation is performed for each working LSP.

#### 2. Without Route Pre-computation (or Full LSP re-routing)

An end-to-end restoration LSP is dynamically established after the failure(s) occur(s). After failure occurrence, one or more (disjoint) paths for the restoration LSP are dynamically computed and



one is selected. As such, one can define this as a complete "LSP re-routing" mechanism.

No reservation or selection of resources is performed along the restoration path before failure occurrence. As a result, there is no guarantee that a restoration LSP is available when a failure occurs.

The expected total restoration time  $T$  is thus equal to  $T_c (+ T_s) + T_{rs}$ . Therefore, time performance between these two approaches differs by the time required for route computation  $T_c$  (and its potential selection time,  $T_s$ ).

### 5.5.3. Pre-Planned LSP Restoration

Pre-planned LSP restoration (also referred to as pre-planned LSP re-routing) implies that the restoration LSP is pre-signaled. This in turn implies the reservation of recovery resources along the restoration path. Two cases can be defined based on whether the recovery resources are pre-selected.

#### 1. With resource reservation and without resource pre-selection

Before failure occurrence, an end-to-end restoration path is pre-selected from a set of pre-computed (disjoint) paths. The restoration LSP is signaled along this pre-selected path to reserve resources at each node, but these resources are not selected.

In this case, the resources reserved for each restoration LSP may be dedicated or shared between multiple restoration LSPs whose working LSPs are not expected to fail simultaneously. Local node policies can be applied to define the degree to which these resources can be shared across independent failures. Also, since a restoration scheme is considered, resource sharing should not be limited to restoration LSPs that start and end at the same ingress and egress nodes. Therefore, each node participating in this scheme is expected to receive some feedback information on the sharing degree of the recovery resource(s) that this scheme involves.

Upon failure detection/notification message reception, signaling is initiated along the restoration path to select the resources, and to perform the appropriate operation at each node crossed by the restoration LSP (e.g., cross-connections). If lower priority LSPs were established using the restoration resources, they must be preempted when the restoration LSP is activated.

Thus, the expected total restoration time  $T$  is equal to  $T_{as}$  (post-failure activation), while operations performed before failure occurrence take  $T_c + T_s + T_r$ .



## 2. With both resource reservation and resource pre-selection

Before failure occurrence, an end-to-end restoration path is pre-selected from a set of pre-computed (disjoint) paths. The restoration LSP is signaled along this pre-selected path to reserve AND select resources at each node, but these resources are not committed at the data plane level. So that the selection of the recovery resources is committed at the control plane level only, no cross-connections are performed along the restoration path.

In this case, the resources reserved and selected for each restoration LSP may be dedicated or even shared between multiple restoration LSPs whose associated working LSPs are not expected to fail simultaneously. Local node policies can be applied to define the degree to which these resources can be shared across independent failures. Also, because a restoration scheme is considered, resource sharing should not be limited to restoration LSPs that start and end at the same ingress and egress nodes. Therefore, each node participating in this scheme is expected to receive some feedback information on the sharing degree of the recovery resource(s) that this scheme involves.

Upon failure detection/notification message reception, signaling is initiated along the restoration path to activate the reserved and selected resources, and to perform the appropriate operation at each node crossed by the restoration LSP (e.g., cross-connections). If lower priority LSPs were established using the restoration resources, they must be preempted when the restoration LSP is activated.

Thus, the expected total restoration time  $T$  is equal to  $T_a$  (post-failure activation), while operations performed before failure occurrence take  $T_c + T_s + T_{rs}$ . Therefore, time performance between these two approaches differs only by the time required for resource selection during the activation of the recovery LSP (i.e.,  $T_s - T_a$ ).

### 5.5.4. LSP Segment Restoration

The above approaches can be applied on an edge-to-edge LSP basis rather than end-to-end LSP basis (i.e., to reduce the global recovery time) by allowing the recovery of the individual LSP segments constituting the end-to-end LSP.

Also, by using the horizontal hierarchy approach described in [Section 7.1](#), an end-to-end LSP can be recovered by multiple recovery mechanisms applied on an LSP segment basis (e.g., 1:1 edge-to-edge LSP protection in a metro network, and M:N edge-to-edge protection in the core). These mechanisms are ideally independent and may even use different failure localization and notification mechanisms.





## **6. Reversion**

Reversion (a.k.a. normalization) is defined as the mechanism allowing switching of normal traffic from the recovery LSP/span to the working LSP/span previously under failure condition. Use of normalization is at the discretion of the recovery domain policy. Normalization may impact the normal traffic (a second hit) depending on the normalization mechanism used.

If normalization is supported, then 1) the LSP/span must be returned to the working LSP/span when the failure condition clears and 2) the capability to de-activate (turn-off) the use of reversion should be provided. De-activation of reversion should not impact the normal traffic, regardless of whether it is currently using the working or recovery LSP/span.

Note: during the failure, the reuse of any non-failed resources (e.g., LSP and/or spans) belonging to the working LSP/span is under the discretion of recovery domain policy.

### **6.1. Wait-To-Restore (WTR)**

A specific mechanism (Wait-To-Restore) is used to prevent frequent recovery switching operations due to an intermittent defect (e.g., Bit Error Rate (BER) fluctuating around the SD threshold).

First, an LSP/span under failure condition must become fault-free, e.g., a BER less than a certain recovery threshold. After the recovered LSP/span (i.e., the previously working LSP/span) meets this criterion, a fixed period of time shall elapse before normal traffic uses the corresponding resources again. This duration called Wait-To-Restore (WTR) period or timer is generally on the order of a few minutes (for instance, 5 minutes) and should be capable of being set. The WTR timer may be either a fixed period, or provide for incrementally longer periods before retrying. An SF or SD condition on the previously working LSP/span will override the WTR timer value (i.e., the WTR cancels and the WTR timer will restart).

### **6.2. Revertive Mode Operation**

In revertive mode of operation, when the recovery LSP/span is no longer required, i.e., the failed working LSP/span is no longer in SD or SF condition, a local Wait-to-Restore (WTR) state will be activated before switching the normal traffic back to the recovered working LSP/span.

During the reversion operation, since this state becomes the highest in priority, signaling must maintain the normal traffic on the



recovery LSP/span from the previously failed working LSP/span. Moreover, during this WTR state, any null traffic or extra traffic (if applicable) request is rejected.

However, deactivation (cancellation) of the wait-to-restore timer may occur if there are higher priority request attempts. That is, the recovery LSP/span usage by the normal traffic may be preempted if a higher priority request for this recovery LSP/span is attempted.

### 6.3. Orphans

When a reversion operation is requested, normal traffic must be switched from the recovery to the recovered working LSP/span. A particular situation occurs when the previously working LSP/span cannot be recovered, so normal traffic cannot be switched back. In that case, the LSP/span under failure condition (also referred to as "orphan") must be cleared (i.e., removed) from the pool of resources allocated for normal traffic. Otherwise, potential de-synchronization between the control and transport plane resource usage can appear. Depending on the signaling protocol capabilities and behavior, different mechanisms are expected here.

Therefore, any reserved or allocated resources for the LSP/span under failure condition must be unreserved/de-allocated. Several ways can be used for that purpose: wait for the clear-out time interval to elapse, initiate a deletion from the ingress or the egress node, or trigger the initiation of deletion from an entity (such as an EMS or NMS) capable of reacting upon reception of an appropriate notification message.

## 7. Hierarchies

Recovery mechanisms are being made available at multiple (if not all) transport layers within so-called "IP/MPLS-over-optical" networks. However, each layer has certain recovery features, and one needs to determine the exact impact of the interaction between the recovery mechanisms provided by these layers.

Hierarchies are used to build scalable complex systems. By hiding the internal details, abstraction is used as a mechanism to build large networks or as a technique for enforcing technology, topological, or administrative boundaries. The same hierarchical concept can be applied to control the network survivability. Network survivability is the set of capabilities that allow a network to restore affected traffic in the event of a failure. Network survivability is defined further in [RFC4427]. In general, it is expected that the recovery action is taken by the recoverable LSP/span closest to the failure in order to avoid the multiplication



of recovery actions. Moreover, recovery hierarchies also can be bound to control plane logical partitions (e.g., administrative or topological boundaries). Each logical partition may apply different recovery mechanisms.

In brief, it is commonly accepted that the lower layers can provide coarse but faster recovery while the higher layers can provide finer but slower recovery. Moreover, it is also desirable to avoid similar layers with functional overlaps in order to optimize network resource utilization and processing overhead, since repeating the same capabilities at each layer does not create any added value for the network as a whole. In addition, even if a lower layer recovery mechanism is enabled, it does not prevent the additional provision of a recovery mechanism at the upper layer. The inverse statement does not necessarily hold; that is, enabling an upper layer recovery mechanism may prevent the use of a lower layer recovery mechanism. In this context, this section analyzes these hierarchical aspects including the physical (passive) layer(s).

### **7.1. Horizontal Hierarchy (Partitioning)**

A horizontal hierarchy is defined when partitioning a single-layer network (and its control plane) into several recovery domains. Within a domain, the recovery scope may extend over a link (or span), LSP segment, or even an end-to-end LSP. Moreover, an administrative domain may consist of a single recovery domain or can be partitioned into several smaller recovery domains. The operator can partition the network into recovery domains based on physical network topology, control plane capabilities, or various traffic engineering constraints.

An example often addressed in the literature is the metro-core-metro application (sometimes extended to a metro-metro/core-core) within a single transport layer (see [Section 7.2](#)). For such a case, an end-to-end LSP is defined between the ingress and egress metro nodes, while LSP segments may be defined within the metro or core sub-networks. Each of these topological structures determines a so-called "recovery domain" since each of the LSPs they carry can have its own recovery type (or even scheme). The support of multiple recovery types and schemes within a sub-network is referred to as a "multi-recovery capable domain" or simply "multi-recovery domain".

### **7.2. Vertical Hierarchy (Layers)**

It is very challenging to combine the different recovery capabilities available across the path (i.e., switching capable) and section layers to ensure that certain network survivability objectives are met for the network-supported services.



As a first analysis step, one can draw the following guidelines for a vertical coordination of the recovery mechanisms:

- The lower the layer, the faster the notification and switching.
- The higher the layer, the finer the granularity of the recoverable entity and therefore the granularity of the recovery resource.

Moreover, in the context of this analysis, a vertical hierarchy consists of multiple layered transport planes providing different:

- Discrete bandwidth granularities for non-packet LSPs such as OCh, ODUk, STS\_SPE/HOVC, and VT\_SPE/LOVC LSPs and continuous bandwidth granularities for packet LSPs.
- Potential recovery capabilities with different temporal granularities: ranging from milliseconds to tens of seconds

Note: based on the bandwidth granularity, we can determine four classes of vertical hierarchies: (1) packet over packet, (2) packet over circuit, (3) circuit over packet, and (4) circuit over circuit. Below we briefly expand on (4) only. (2) is covered in [[RFC3386](#)]. (1) is extensively covered by the MPLS Working Group, and (3) by the PWE3 Working Group.

In SONET/SDH environments, one typically considers the VT\_SPE/LOVC and STS\_SPE/HOVC as independent layers (for example, VT\_SPE/LOVC LSP uses the underlying STS\_SPE/HOVC LSPs as links). In OTN, the ODUk path layers will lie on the OCh path layer, i.e., the ODUk LSPs use the underlying OCh LSPs as OTUk links. Note here that lower layer LSPs may simply be provisioned and not necessarily dynamically triggered or established (control driven approach). In this context, an LSP at the path layer (i.e., established using GMPLS signaling), such as an optical channel LSP, appears at the OTUk layer as a link, controlled by a link management protocol such as LMP.

The first key issue with multi-layer recovery is that achieving individual or bulk LSP recovery will be as efficient as the underlying link (local span) recovery. In such a case, the span can be either protected or unprotected, but the LSP it carries must be (at least locally) recoverable. Therefore, the span recovery process can be either independent when protected (or restorable), or triggered by the upper LSP recovery process. The former case requires coordination to achieve subsequent LSP recovery. Therefore, in order to achieve robustness and fast convergence, multi-layer recovery requires a fine-tuned coordination mechanism.





Moreover, in the absence of adequate recovery mechanism coordination (for instance, a pre-determined coordination when using a hold-off timer), a failure notification may propagate from one layer to the next one within a recovery hierarchy. This can cause "collisions" and trigger simultaneous recovery actions that may lead to race conditions and, in turn, reduce the optimization of the resource utilization and/or generate global instabilities in the network (see [\[MANCHESTER\]](#)). Therefore, a consistent and efficient escalation strategy is needed to coordinate recovery across several layers.

One can expect that the definition of the recovery mechanisms and protocol(s) is technology-independent so that they can be consistently implemented at different layers; this would in turn simplify their global coordination. Moreover, as mentioned in [\[RFC3386\]](#), some looser form of coordination and communication between (vertical) layers such as a consistent hold-off timer configuration (and setup through signaling during the working LSP establishment) can be considered, thereby allowing the synchronization between recovery actions performed across these layers.

#### **[7.2.1.](#) Recovery Granularity**

In most environments, the design of the network and the vertical distribution of the LSP bandwidth are such that the recovery granularity is finer at higher layers. The OTN and SONET/SDH layers can recover only the whole section or the individual connections they transports whereas the IP/MPLS control plane can recover individual packet LSPs or groups of packet LSPs independently of their granularity. On the other side, the recovery granularity at the sub-wavelength level (i.e., SONET/SDH) can be provided only when the network includes devices switching at the same granularity (and thus not with optical channel level). Therefore, the network layer can deliver control-plane-driven recovery mechanisms on a per-LSP basis if and only if these LSPs have their corresponding switching granularity supported at the transport plane level.

#### **[7.3.](#) Escalation Strategies**

There are two types of escalation strategies (see [\[DEMEESTER\]](#)): bottom-up and top-down.

The bottom-up approach assumes that lower layer recovery types and schemes are more expedient and faster than upper layer ones. Therefore, we can inhibit or hold off higher layer recovery. However, this assumption is not entirely true. Consider for instance a SONET/SDH based protection mechanism (with a protection switching time of less than 50 ms) lying on top of an OTN restoration mechanism (with a restoration time of less than 200 ms). Therefore, this



assumption should be (at least) clarified as: the lower layer recovery mechanism is expected to be faster than the upper level one, if the same type of recovery mechanism is used at each layer.

Consequently, taking into account the recovery actions at the different layers in a bottom-up approach: if lower layer recovery mechanisms are provided and sequentially activated in conjunction with higher layer ones, the lower layers must have an opportunity to recover normal traffic before the higher layers do. However, if lower layer recovery is slower than higher layer recovery, the lower layer must either communicate the failure-related information to the higher layer(s) (and allow it to perform recovery), or use a hold-off timer in order to temporarily set the higher layer recovery action in a "standby mode". Note that the a priori information exchange between layers concerning their efficiency is not within the current scope of this document. Nevertheless, the coordination functionality between layers must be configurable and tunable.

For example, coordination between the optical and packet layer control plane enables the optical layer to perform the failure management operations (in particular, failure detection and notification) while giving to the packet layer control plane the authority to decide and perform the recovery actions. If the packet layer recovery action is unsuccessful, fallback at the optical layer can be performed subsequently.

The top-down approach attempts service recovery at the higher layers before invoking lower layer recovery. Higher layer recovery is service selective, and permits "per-CoS" or "per-connection" re-routing. With this approach, the most important aspect is that the upper layer should provide its own reliable and independent failure detection mechanism from the lower layer.

[DEMEESTER] also suggests recovery mechanisms incorporating a coordinated effort shared by two adjacent layers with periodic status updates. Moreover, some of these recovery operations can be pre-assigned (on a per-link basis) to a certain layer, e.g., a given link will be recovered at the packet layer while another will be recovered at the optical layer.

#### **7.4. Disjointness**

Having link and node diverse working and recovery LSPs/spans does not guarantee their complete disjointness. Due to the common physical layer topology (passive), additional hierarchical concepts, such as the Shared Risk Link Group (SRLG), and mechanisms, such as SRLG diverse path computation, must be developed to provide complete working and recovery LSP/span disjointness (see [[IPO-IMP](#)] and



[RFC4202]). Otherwise, a failure affecting the working LSP/span would also potentially affect the recovery LSP/span; one refers to such an event as "common failure".

#### 7.4.1. SRLG Disjointness

A Shared Risk Link Group (SRLG) is defined as the set of links sharing a common risk (such as a common physical resource such as a fiber link or a fiber cable). For instance, a set of links  $L$  belongs to the same SRLG  $s$ , if they are provisioned over the same fiber link  $f$ .

The SRLG properties can be summarized as follows:

- 1) A link belongs to more than one SRLG if and only if it crosses one of the resources covered by each of them.
- 2) Two links belonging to the same SRLG can belong individually to (one or more) other SRLGs.
- 3) The SRLG set  $S$  of an LSP is defined as the union of the individual SRLG  $s$  of the individual links composing this LSP.

SRLG disjointness is also applicable to LSPs:

The LSP SRLG disjointness concept is based on the following postulate: an LSP (i.e., a sequence of links and nodes) covers an SRLG if and only if it crosses one of the links or nodes belonging to that SRLG.

Therefore, the SRLG disjointness for LSPs, can be defined as follows: two LSPs are disjoint with respect to an SRLG  $s$  if and only if they do not cover simultaneously this SRLG  $s$ .

Whilst the SRLG disjointness for LSPs with respect to a set  $S$  of SRLGs, is defined as follows: two LSPs are disjoint with respect to a set of SRLGs  $S$  if and only if the set of SRLGs that are common to both LSPs is disjoint from set  $S$ .

The impact on recovery is noticeable: SRLG disjointness is a necessary (but not a sufficient) condition to ensure network survivability. With respect to the physical network resources, a working-recovery LSP/span pair must be SRLG-disjoint in case of dedicated recovery type. On the other hand, in case of shared recovery, a group of working LSP/spans must be mutually SRLG-disjoint in order to allow for a (single and common) shared recovery LSP that is itself SRLG-disjoint from each of the working LSPs/spans.



## **8. Recovery Mechanisms Analysis**

In order to provide a structured analysis of the recovery mechanisms detailed in the previous sections, the following dimensions can be considered:

1. Fast convergence (performance): provide a mechanism that aggregates multiple failures (implying fast failure detection and correlation mechanisms) and fast recovery decision independently of the number of failures occurring in the optical network (also implying a fast failure notification).
2. Efficiency (scalability): minimize the switching time required for LSP/span recovery independently of the number of LSPs/spans being recovered (this implies efficient failure correlation, fast failure notification, and time-efficient recovery mechanisms).
3. Robustness (availability): minimize the LSP/span downtime independently of the underlying topology of the transport plane (this implies a highly responsive recovery mechanism).
4. Resource optimization (optimality): minimize the resource capacity, including LSPs/spans and nodes (switching capacity), required for recovery purposes; this dimension can also be referred to as optimizing the sharing degree of the recovery resources.
5. Cost optimization: provide a cost-effective recovery type/scheme.

However, these dimensions are either outside the scope of this document (such as cost optimization and recovery path computational aspects) or mutually conflicting. For instance, it is obvious that providing a 1+1 LSP protection minimizes the LSP downtime (in case of failure) while being non-scalable and consuming recovery resource without enabling any extra-traffic.

The following sections analyze the recovery phases and mechanisms detailed in the previous sections with respect to the dimensions described above in order to assess the GMPLS protocol suite capabilities and applicability. In turn, this allows the evaluation of the potential need for further GMPLS signaling and routing extensions.





### **8.1. Fast Convergence (Detection/Correlation and Hold-off Time)**

Fast convergence is related to the failure management operations. It refers to the time elapsed between failure detection/correlation and hold-off time, the point at which the recovery switching actions are initiated. This point has been detailed in [Section 4](#).

### **8.2. Efficiency (Recovery Switching Time)**

In general, the more pre-assignment/pre-planning of the recovery LSP/span, the more rapid the recovery is. Because protection implies pre-assignment (and cross-connection) of the protection resources, in general, protection recovers faster than restoration.

Span restoration is likely to be slower than most span protection types; however this greatly depends on the efficiency of the span restoration signaling. LSP restoration with pre-sigaled and pre-selected recovery resources is likely to be faster than fully dynamic LSP restoration, especially because of the elimination of any potential crankback during the recovery LSP establishment.

If one excludes the crankback issue, the difference between dynamic and pre-planned restoration depends on the restoration path computation and selection time. Since computational considerations are outside the scope of this document, it is up to the vendor to determine the average and maximum path computation time in different scenarios and to the operator to decide whether or not dynamic restoration is advantageous over pre-planned schemes that depend on the network environment. This difference also depends on the flexibility provided by pre-planned restoration versus dynamic restoration. Pre-planned restoration implies a somewhat limited number of failure scenarios (that can be due, for instance, to local storage capacity limitation). Dynamic restoration enables on-demand path computation based on the information received through failure notification message, and as such, it is more robust with respect to the failure scenario scope.

Moreover, LSP segment restoration, in particular, dynamic restoration (i.e., no path pre-computation, so none of the recovery resource is pre-reserved) will generally be faster than end-to-end LSP restoration. However, local LSP restoration assumes that each LSP segment end-point has enough computational capacity to perform this operation while end-to-end LSP restoration requires only that LSP end-points provide this path computation capability.

Recovery time objectives for SONET/SDH protection switching (not including time to detect failure) are specified in [\[G.841\]](#) at 50 ms, taking into account constraints on distance, number of connections



involved, and in the case of ring enhanced protection, number of nodes in the ring. Recovery time objectives for restoration mechanisms have been proposed through a separate effort [[RFC3386](#)].

### 8.3. Robustness

In general, the less pre-assignment (protection)/pre-planning (restoration) of the recovery LSP/span, the more robust the recovery type or scheme is to a variety of single failures, provided that adequate resources are available. Moreover, the pre-selection of the recovery resources gives (in the case of multiple failure scenarios) less flexibility than no recovery resource pre-selection. For instance, if failures occur that affect two LSPs sharing a common link along their restoration paths, then only one of these LSPs can be recovered. This occurs unless the restoration path of at least one of these LSPs is re-computed, or the local resource assignment is modified on the fly.

In addition, recovery types and schemes with pre-planned recovery resources (in particular, LSP/spans for protection and LSPs for restoration purposes) will not be able to recover from failures that simultaneously affect both the working and recovery LSP/span. Thus, the recovery resources should ideally be as disjoint as possible (with respect to link, node, and SRLG) from the working ones, so that any single failure event will not affect both working and recovery LSP/span. In brief, working and recovery resources must be fully diverse in order to guarantee that a given failure will not affect simultaneously the working and the recovery LSP/span. Also, the risk of simultaneous failure of the working and the recovery LSPs can be reduced. It is reduced by computing a new recovery path whenever a failure occurs along one of the recovery LSPs or by computing a new recovery path and provision the corresponding LSP whenever a failure occurs along a working LSP/span. Both methods enable the network to maintain the number of available recovery path constant.

The robustness of a recovery scheme is also determined by the amount of pre-reserved (i.e., signaled) recovery resources within a given shared resource pool: as the sharing degree of recovery resources increases, the recovery scheme becomes less robust to multiple LSP/span failure occurrences. Recovery schemes, in particular restoration, with pre-sigaled resource reservation (with or without pre-selection) should be capable of reserving an adequate amount of resource to ensure recovery from any specific set of failure events, such as any single SRLG failure, any two SRLG failures, etc.



#### 8.4. Resource Optimization

It is commonly admitted that sharing recovery resources provides network resource optimization. Therefore, from a resource utilization perspective, protection schemes are often classified with respect to their degree of sharing recovery resources with the working entities. Moreover, non-permanent bridging protection types allow (under normal conditions) for extra-traffic over the recovery resources.

From this perspective, the following statements are true:

- 1) 1+1 LSP/Span protection is the most resource-consuming protection type because it does not allow for any extra traffic.
- 2) 1:1 LSP/span recovery requires dedicated recovery LSP/span allowing for extra traffic.
- 3) 1:N and M:N LSP/span recovery require 1 (and M, respectively) recovery LSP/span (shared between the N working LSP/span) allowing for extra traffic.

Obviously, 1+1 protection precludes, and 1:1 recovery does not allow for any recovery LSP/span sharing, whereas 1:N and M:N recovery do allow sharing of 1 (M, respectively) recovery LSP/spans between N working LSP/spans. However, despite the fact that 1:1 LSP recovery precludes the sharing of the recovery LSP, the recovery schemes that can be built from it (e.g.,  $(1:1)^N$ , see [Section 5.4](#)) do allow sharing of its recovery resources. In addition, the flexibility in the usage of shared recovery resources (in particular, shared links) may be limited because of network topology restrictions, e.g., fixed ring topology for traditional enhanced protection schemes.

On the other hand, when using LSP restoration with pre-sigaled resource reservation, the amount of reserved restoration capacity is determined by the local bandwidth reservation policies. In LSP restoration schemes with re-provisioning, a pool of spare resources can be defined from which all resources are selected after failure occurrence for the purpose of restoration path computation. The degree to which restoration schemes allow sharing amongst multiple independent failures is then directly inferred from the size of the resource pool. Moreover, in all restoration schemes, spare resources can be used to carry preemptible traffic (thus over preemptible LSP/span) when the corresponding resources have not been committed for LSP/span recovery purposes.

From this, it clearly follows that less recovery resources (i.e., LSP/spans and switching capacity) have to be allocated to a shared



recovery resource pool if a greater sharing degree is allowed. Thus, the network survivability level is determined by the policy that defines the amount of shared recovery resources and by the maximum sharing degree allowed for these recovery resources.

#### 8.4.1. Recovery Resource Sharing

When recovery resources are shared over several LSP/Spans, the use of the Maximum Reservable Bandwidth, the Unreserved Bandwidth, and the Maximum LSP Bandwidth (see [RFC4202]) provides the information needed to obtain the optimization of the network resources allocated for shared recovery purposes.

The Maximum Reservable Bandwidth is defined as the Maximum Link Bandwidth but it may be greater in case of link over-subscription.

The Unreserved Bandwidth (at priority  $p$ ) is defined as the bandwidth not yet reserved on a given TE link (its initial value for each priority  $p$  corresponds to the Maximum Reservable Bandwidth). Last, the Maximum LSP Bandwidth (at priority  $p$ ) is defined as the smaller of Unreserved Bandwidth (at priority  $p$ ) and Maximum Link Bandwidth.

Here, one generally considers a recovery resource sharing degree (or ratio) to globally optimize the shared recovery resource usage. The distribution of the bandwidth utilization per TE link can be inferred from the per-priority bandwidth pre-allocation. By using the Maximum LSP Bandwidth and the Maximum Reservable Bandwidth, the amount of (over-provisioned) resources that can be used for shared recovery purposes is known from the IGP.

In order to analyze this behavior, we define the difference between the Maximum Reservable Bandwidth (in the present case, this value is greater than the Maximum Link Bandwidth) and the Maximum LSP Bandwidth per TE link  $i$  as the Maximum Shareable Bandwidth or  $\max\_R[i]$ . Within this quantity, the amount of bandwidth currently allocated for shared recovery per TE link  $i$  is defined as  $R[i]$ . Both quantities are expressed in terms of discrete bandwidth units (and thus, the Minimum LSP Bandwidth is of one bandwidth unit).

The knowledge of this information available per TE link can be exploited in order to optimize the usage of the resources allocated per TE link for shared recovery. If one refers to  $r[i]$  as the actual bandwidth per TE link  $i$  (in terms of discrete bandwidth units) committed for shared recovery, then the following quantity must be maximized over the potential TE link candidates:

$$\sum_{i=1}^N [(R[i] - r[i]) / (t[i] - b[i])]$$



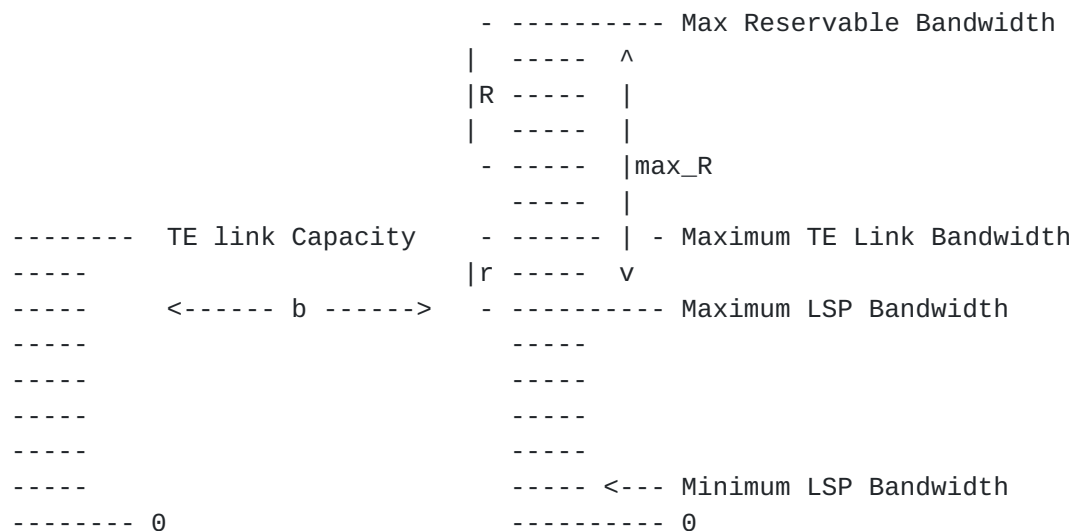


or equivalently:  $\sum_{i=1}^N [(R\{i\} - r\{i\})/r\{i\}]$

with  $R\{i\} \geq 1$  and  $r\{i\} \geq 1$  (in terms of per component bandwidth unit)

In this formula,  $N$  is the total number of links traversed by a given LSP,  $t[i]$  the Maximum Link Bandwidth per TE link  $i$ , and  $b[i]$  the sum per TE link  $i$  of the bandwidth committed for working LSPs and other recovery LSPs (thus except "shared bandwidth" LSPs). The quantity  $[(R\{i\} - r\{i\})/r\{i\}]$  is defined as the Shared (Recovery) Bandwidth Ratio per TE link  $i$ . In addition, TE links for which  $R[i]$  reaches  $\max\_R[i]$  or for which  $r[i] = 0$  are pruned during shared recovery path computation as well as TE links for which  $\max\_R[i] = r[i]$  that can simply not be shared.

More generally, one can draw the following mapping between the available bandwidth at the transport and control plane level:



Note that the above approach does not require the flooding of any per LSP information or any detailed distribution of the bandwidth allocation per component link or individual ports or even any per-priority shareable recovery bandwidth information (using a dedicated sub-TLV). The latter would provide the same capability as the already defined Maximum LSP bandwidth per-priority information. This approach is referred to as a Partial (or Aggregated) Information Routing as described in [[KODIALAM1](#)] and [[KODIALAM2](#)]. They show that the difference obtained with a Full (or Complete) Information Routing approach (where for the whole set of working and recovery LSPs, the amount of bandwidth units they use per-link is known at each node and for each link) is clearly negligible. The Full Information Routing



approach is detailed in [GLI]. Note also that both approaches rely on the deterministic knowledge (at different degrees) of the network topology and resource usage status.

Moreover, extending the GMPLS signaling capabilities can enhance the Partial Information Routing approach. It is enhanced by allowing working-LSP-related information and, in particular, its path (including link and node identifiers) to be exchanged with the recovery LSP request. This enables more efficient admission control at upstream nodes of shared recovery resources, and in particular, links (see [Section 8.4.3](#)).

#### **8.4.2. Recovery Resource Sharing and SRLG Recovery**

Resource shareability can also be maximized with respect to the number of times each SRLG is protected by a recovery resource (in particular, a shared TE link) and methods can be considered for avoiding contention of the shared recovery resources in case of single SRLG failure. These methods enable the sharing of recovery resources between two (or more) recovery LSPs, if their respective working LSPs are mutually disjoint with respect to link, node, and SRLGs. Then, a single failure does not simultaneously disrupt several (or at least two) working LSPs.

For instance, [BOUILLET] shows that the Partial Information Routing approach can be extended to cover recovery resource shareability with respect to SRLG recoverability (i.e., the number of times each SRLG is recoverable). By flooding this aggregated information per TE link, path computation and selection of SRLG-diverse recovery LSPs can be optimized with respect to the sharing of recovery resource reserved on each TE link. This yields a performance difference of less than 5%, which is negligible compared to the corresponding Full Information Flooding approach (see [GLI]).

For this purpose, additional extensions to [RFC4202] in support of path computation for shared mesh recovery have been often considered in the literature. TE link attributes would include, among others, the current number of recovery LSPs sharing the recovery resources reserved on the TE link, and the current number of SRLGs recoverable by this amount of (shared) recovery resources reserved on the TE link. The latter is equivalent to the current number of SRLGs that will be recovered by the recovery LSPs sharing the recovery resource reserved on the TE link. Then, if explicit SRLG recoverability is considered, a TE link attribute would be added that includes the explicit list of SRLGs (recoverable by the shared recovery resource reserved on the TE link) and their respective shareable recovery bandwidths. The latter information is equivalent to the shareable recovery bandwidth per SRLG (or per group of SRLGs), which implies

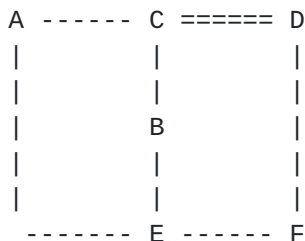


that the amount of shareable bandwidth and the number of listed SRLGs will decrease over time.

Compared to the case of recovery resource sharing only (regardless of SRLG recoverability, as described in [Section 8.4.1](#)), these additional TE link attributes would potentially deliver better path computation and selection (at a distinct ingress node) for shared mesh recovery purposes. However, due to the lack of evidence of better efficiency and due to the complexity that such extensions would generate, they are not further considered in the scope of the present analysis. For instance, a per-SRLG group minimum/maximum shareable recovery bandwidth is restricted by the length that the corresponding (sub-) TLV may take and thus the number of SRLGs that it can include. Therefore, the corresponding parameter should not be translated into GMPLS routing (or even signaling) protocol extensions in the form of TE link sub-TLV.

#### **8.4.3. Recovery Resource Sharing, SRLG Disjointness and Admission Control**

Admission control is a strict requirement to be fulfilled by nodes giving access to shared links. This can be illustrated using the following network topology:



Node A creates a working LSP to D (A-C-D), B creates simultaneously a working LSP to D (B-C-D) and a recovery LSP (B-E-F-D) to the same destination. Then, A decides to create a recovery LSP to D (A-E-F-D), but since the C-D span carries both working LSPs, node E should either assign a dedicated resource for this recovery LSP or reject this request if the C-D span has already reached its maximum recovery bandwidth sharing ratio. In the latter case, C-D span failure would imply that one of the working LSP would not be recoverable.

Consequently, node E must have the required information to perform admission control for the recovery LSP requests it processes (implying for instance, that the path followed by the working LSP is carried with the corresponding recovery LSP request). If node E can guarantee that the working LSPs (A-C-D and B-C-D) are SRLG disjoint over the C-D span, it may securely accept the incoming recovery LSP request and assign to the recovery LSPs (A-E-F-D and B-E-F-D) the



same resources on the link E-F. This may occur if the link E-F has not yet reached its maximum recovery bandwidth sharing ratio. In this example, one assumes that the node failure probability is negligible compared to the link failure probability.

To achieve this, the path followed by the working LSP is transported with the recovery LSP request and examined at each upstream node of potentially shareable links. Admission control is performed using the interface identifiers (included in the path) to retrieve in the TE DataBase the list of SRLG IDs associated to each of the working LSP links. If the working LSPs (A-C-D and B-C-D) have one or more link or SRLG ID in common (in this example, one or more SRLG id in common over the span C-D), node E should not assign the same resource over link E-F to the recovery LSPs (A-E-F-D and B-E-F-D). Otherwise, one of these working LSPs would not be recoverable if C-D span failure occurred.

There are some issues related to this method; the major one is the number of SRLG IDs that a single link can cover (more than 100, in complex environments). Moreover, when using link bundles, this approach may generate the rejection of some recovery LSP requests. This occurs when the SRLG sub-TLV corresponding to a link bundle includes the union of the SRLG id list of all the component links belonging to this bundle (see [[RFC4202](#)] and [[RFC4201](#)]).

In order to overcome this specific issue, an additional mechanism may consist of querying the nodes where the information would be available (in this case, node E would query C). The main drawback of this method is that (in addition to the dedicated mechanism(s) it requires) it may become complex when several common nodes are traversed by the working LSPs. Therefore, when using link bundles, solving this issue is closely related to the sequence of the recovery operations. Per-component flooding of SRLG identifiers would deeply impact the scalability of the link state routing protocol. Therefore, one may rely on the usage of an on-line accessible network management system.





## 9. Summary and Conclusions

The following table summarizes the different recovery types and schemes analyzed throughout this document.

		Path Search (computation and selection)	
		Pre-planned (a)	Dynamic (b)
Path Setup	1	faster recovery	Does not apply
		less flexible	
		less robust	
		most resource-consuming	
	2	relatively fast recovery	Does not apply
		relatively flexible	
		relatively robust	
		resource consumption	
		depends on sharing degree	
	3	relatively fast recovery	less faster (computation)
		more flexible	most flexible
		relatively robust	most robust
		less resource-consuming	least resource-consuming
		depends on sharing degree	

- 1a. Recovery LSP setup (before failure occurrence) with resource reservation (i.e., signaling) and selection is referred to as LSP protection.
- 2a. Recovery LSP setup (before failure occurrence) with resource reservation (i.e., signaling) and with resource pre-selection is referred to as pre-planned LSP re-routing with resource pre-selection. This implies only recovery LSP activation after failure occurrence.
- 3a. Recovery LSP setup (before failure occurrence) with resource reservation (i.e., signaling) and without resource selection is referred to as pre-planned LSP re-routing without resource pre-selection. This implies recovery LSP activation and resource (i.e., label) selection after failure occurrence.
- 3b. Recovery LSP setup after failure occurrence is referred to as LSP re-routing, which is full when recovery LSP path computation occurs after failure occurrence.



Thus, the term pre-planned refers to recovery LSP path pre-computation, signaling (reservation), and a priori resource selection (optional), but not cross-connection. Also, the shared-mesh recovery scheme can be viewed as a particular case of 2a) and 3a), using the additional constraint described in [Section 8.4.3](#).

The implementation of these recovery mechanisms requires only considering extensions to GMPLS signaling protocols (i.e., [[RFC3471](#)] and [[RFC3473](#)]). These GMPLS signaling extensions should mainly focus in delivering (1) recovery LSP pre-provisioning for the cases 1a, 2a, and 3a, (2) LSP failure notification, (3) recovery LSP switching action(s), and (4) reversion mechanisms.

Moreover, the present analysis (see [Section 8](#)) shows that no GMPLS routing extensions are expected to efficiently implement any of these recovery types and schemes.

## **[10.](#) Security Considerations**

This document does not introduce any additional security issue or imply any specific security consideration from [[RFC3945](#)] to the current RSVP-TE GMPLS signaling, routing protocols (OSPF-TE, IS-IS-TE) or network management protocols.

However, the authorization of requests for resources by GMPLS-capable nodes should determine whether a given party, presumably already authenticated, has a right to access the requested resources. This determination is typically a matter of local policy control, for example, by setting limits on the total bandwidth made available to some party in the presence of resource contention. Such policies may become quite complex as the number of users, types of resources, and sophistication of authorization rules increases. This is particularly the case for recovery schemes that assume pre-planned sharing of recovery resources, or contention for resources in case of dynamic re-routing.

Therefore, control elements should match the requests against the local authorization policy. These control elements must be capable of making decisions based on the identity of the requester, as verified cryptographically and/or topologically.

## **[11.](#) Acknowledgements**

The authors would like to thank Fabrice Poppe (Alcatel) and Bart Rousseau (Alcatel) for their revision effort, and Richard Rabbat (Fujitsu Labs), David Griffith (NIST), and Lyndon Ong (Ciena) for their useful comments.



Thanks also to Adrian Farrel for the thorough review of the document.

## **[12.](#) References**

### **[12.1.](#) 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.
- [RFC3473] Berger, L., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions", [RFC 3473](#), January 2003.
- [RFC3945] Mannie, E., "Generalized Multi-Protocol Label Switching (GMPLS) Architecture", [RFC 3945](#), October 2004.
- [RFC4201] Kompella, K., Rekhter, Y., and L. Berger, "Link Bundling in MPLS Traffic Engineering (TE)", [RFC 4201](#), October 2005.
- [RFC4202] Kompella, K., Ed. and Y. Rekhter, Ed., "Routing Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", [RFC 4202](#), October 2005.
- [RFC4204] Lang, J., Ed., "Link Management Protocol (LMP)", [RFC 4204](#), October 2005.
- [RFC4209] Fredette, A., Ed. and J. Lang, Ed., "Link Management Protocol (LMP) for Dense Wavelength Division Multiplexing (DWDM) Optical Line Systems", [RFC 4209](#), October 2005.
- [RFC4427] Mannie E., Ed. and D. Papadimitriou, Ed., "Recovery (Protection and Restoration) Terminology for Generalized Multi-Protocol Label Switching (GMPLS)", [RFC 4427](#), March 2006.

### **[12.2.](#) Informative References**

- [BOUILLET] E. Bouillet, et al., "Stochastic Approaches to Compute Shared Meshed Restored Lightpaths in Optical Network Architectures," IEEE Infocom 2002, New York City, June 2002.



- [DEMEESTER] P. Demeester, et al., "Resilience in Multilayer Networks," IEEE Communications Magazine, Vol. 37, No. 8, pp. 70-76, August 1998.
- [GLI] G. Li, et al., "Efficient Distributed Path Selection for Shared Restoration Connections," IEEE Infocom 2002, New York City, June 2002.
- [IPO-IMP] Strand, J. and A. Chiu, "Impairments and Other Constraints on Optical Layer Routing", [RFC 4054](#), May 2005.
- [KODIALAM1] M. Kodialam and T.V. Lakshman, "Restorable Dynamic Quality of Service Routing," IEEE Communications Magazine, pp. 72-81, June 2002.
- [KODIALAM2] M. Kodialam and T.V. Lakshman, "Dynamic Routing of Restorable Bandwidth-Guaranteed Tunnels using Aggregated Network Resource Usage Information," IEEE/ ACM Transactions on Networking, pp. 399-410, June 2003.
- [MANCHESTER] J. Manchester, P. Bonenfant and C. Newton, "The Evolution of Transport Network Survivability," IEEE Communications Magazine, August 1999.
- [RFC3386] Lai, W. and D. McDysan, "Network Hierarchy and Multilayer Survivability", [RFC 3386](#), November 2002.
- [T1.105] ANSI, "Synchronous Optical Network (SONET): Basic Description Including Multiplex Structure, Rates, and Formats," ANSI T1.105, January 2001.
- [WANG] J. Wang, L. Sahasrabudhe, and B. Mukherjee, "Path vs. Subpath vs. Link Restoration for Fault Management in IP-over-WDM Networks: Performance Comparisons Using GMPLS Control Signaling," IEEE Communications Magazine, pp. 80-87, November 2002.

For information on the availability of the following documents, please see <http://www.itu.int>

- [G.707] ITU-T, "Network Node Interface for the Synchronous Digital Hierarchy (SDH)," Recommendation G.707, October 2000.
- [G.709] ITU-T, "Network Node Interface for the Optical Transport Network (OTN)," Recommendation G.709, February 2001 (and Amendment no.1, October 2001).





- [G.783] ITU-T, "Characteristics of Synchronous Digital Hierarchy (SDH) Equipment Functional Blocks," Recommendation G.783, October 2000.
- [G.798] ITU-T, "Characteristics of optical transport network hierarchy equipment functional block," Recommendation G.798, June 2004.
- [G.806] ITU-T, "Characteristics of Transport Equipment - Description Methodology and Generic Functionality", Recommendation G.806, October 2000.
- [G.841] ITU-T, "Types and Characteristics of SDH Network Protection Architectures," Recommendation G.841, October 1998.
- [G.842] ITU-T, "Interworking of SDH network protection architectures," Recommendation G.842, October 1998.
- [G.874] ITU-T, "Management aspects of the optical transport network element," Recommendation G.874, November 2001.

#### Editors' Addresses

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

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

Eric Mannie  
Perceval  
Rue Tenbosch, 9  
1000 Brussels  
Belgium

Phone: +32-2-6409194  
EMail: eric.mannie@perceval.net



## Full Copyright Statement

Copyright (C) The Internet Society (2006).

This document is subject to the rights, licenses and restrictions contained in [BCP 78](#), and except as set forth therein, the authors retain all their rights.

This document and the information contained herein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

## Intellectual Property

The IETF takes no position regarding the validity or scope of any Intellectual Property Rights or other rights that might be claimed to pertain to the implementation or use of the technology described in this document or the extent to which any license under such rights might or might not be available; nor does it represent that it has made any independent effort to identify any such rights. Information on the procedures with respect to rights in RFC documents can be found in [BCP 78](#) and [BCP 79](#).

Copies of IPR disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this specification can be obtained from the IETF on-line IPR repository at <http://www.ietf.org/ipr>.

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights that may cover technology that may be required to implement this standard. Please address the information to the IETF at [ietf-ipr@ietf.org](mailto:ietf-ipr@ietf.org).

## Acknowledgement

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

