Internet Engineering Task Force (IETF)
Request for Comments: 6372
Category: Informational
ISSN: 2070-1721

N. Sprecher, Ed.
Nokia Siemens Networks
A. Farrel, Ed.
Juniper Networks

September 2011

MPLS Transport Profile (MPLS-TP) Survivability Framework

Abstract

Network survivability is the ability of a network to recover traffic delivery following failure or degradation of network resources. Survivability is critical for the delivery of guaranteed network services, such as those subject to strict Service Level Agreements (SLAs) that place maximum bounds on the length of time that services may be degraded or unavailable.

The Transport Profile of Multiprotocol Label Switching (MPLS-TP) is a packet-based transport technology based on the MPLS data plane that reuses many aspects of the MPLS management and control planes.

This document comprises a framework for the provision of survivability in an MPLS-TP network; it describes recovery elements, types, methods, and topological considerations. To enable data-plane recovery, survivability may be supported by the control plane, management plane, and by Operations, Administration, and Maintenance (OAM) functions. This document describes mechanisms for recovering MPLS-TP Label Switched Paths (LSPs). A detailed description of pseudowire recovery in MPLS-TP networks is beyond the scope of this document.

This document is a product of a joint Internet Engineering Task Force (IETF) / International Telecommunication Union Telecommunication Standardization Sector (ITU-T) effort to include an MPLS Transport Profile within the IETF MPLS and Pseudowire Emulation Edge-to-Edge (PWE3) architectures to support the capabilities and functionalities of a packet-based transport network as defined by the ITU-T.

Status of This Memo

This document is not an Internet Standards Track specification; it is published for informational purposes.

This document is a product of the Internet Engineering Task Force (IETF). It represents the consensus of the IETF community. It has received public review and has been approved for publication by the Internet Engineering Steering Group (IESG). Not all documents

approved by the IESG are a candidate for any level of Internet Standard; see Section 2 of RFC 5741.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at http://www.rfc-editor.org/info/rfc6372.

Table of Contents

1. Introduction

Network survivability is the network's ability to recover traffic delivery following the failure or degradation of traffic delivery caused by a network fault or a denial-of-service attack on the network. Survivability plays a critical role in the delivery of reliable services in transport networks. Guaranteed services in the form of Service Level Agreements (SLAs) require a resilient network that very rapidly detects facility or node degradation or failures, and immediately starts to recover network operations in accordance with the terms of the SLA.

The MPLS Transport Profile (MPLS-TP) is described in [RFC5921]. MPLS-TP is designed to be consistent with existing transport network operations and management models, while providing survivability mechanisms, such as protection and restoration. The functionality provided is intended to be similar to or better than that found in established transport networks that set a high benchmark for reliability. That is, it is intended to provide the operator with functions with which they are familiar through their experience with other transport networks, although this does not preclude additional techniques.

This document provides a framework for MPLS-TP-based survivability that meets the recovery requirements specified in [RFC5654]. It uses the recovery terminology defined in [RFC4427], which draws heavily on [G.808.1], and it refers to the requirements specified in [RFC5654].

This document is a product of a joint Internet Engineering Task Force (IETF) / International Telecommunication Union Telecommunication Standardization Sector (ITU-T) effort to include an MPLS Transport Profile within the IETF MPLS and PWE3 architectures to support the capabilities and functionalities of a packet-based transport network, as defined by the ITU-T.

1.1. Recovery Schemes

Various recovery schemes (for protection and restoration) and processes have been defined and analyzed in [RFC4427] and [RFC4428]. These schemes can also be applied in MPLS-TP networks to re-establish end-to-end traffic delivery according to the agreed service parameters, and to trigger recovery from "failed" or "degraded" transport entities. In the context of this document, transport entities are nodes, links, transport path segments, concatenated transport path segments, and entire transport paths. Recovery actions are initiated by the detection of a defect, or by an external request (e.g., an operator's request for manual control of protection switching).

[RFC4427] makes a distinction between protection switching and restoration mechanisms.

- Protection switching uses pre-assigned capacity between nodes, where the simplest scheme has a single, dedicated protection entity for each working entity, while the most complex scheme has m protection entities shared between n working entities (m:n).

- Restoration uses any capacity available between nodes and usually involves rerouting. The resources used for restoration may be pre-planned (i.e., predetermined, but not yet allocated to the recovery path), and recovery priority may be used as a differentiation mechanism to determine which services are recovered and which are not recovered.

Both protection switching and restoration may be either unidirectional or bidirectional; unidirectional implies that protection switching is performed independently for each direction of a bidirectional transport path, while bidirectional means that both directions are switched simultaneously using appropriate coordination, even if the fault applies to only one direction of the path.

Both protection and restoration mechanisms may be either revertive or non-revertive as described in Section 4.11 of [RFC4427].

Preemption priority may be used to determine which services are sacrificed to enable the recovery of other services. Restoration may also be either unidirectional or bidirectional. In general, protection actions are completed within time frames amounting to tens of milliseconds, while automated restoration actions are normally completed within periods ranging from hundreds of milliseconds to a maximum of a few seconds. Restoration is not guaranteed (for example, because network resources may not be available at the time of the defect).

1.2. Recovery Action Initiation

The recovery schemes described in [RFC4427] and evaluated in [RFC4428] are presented in the context of control-plane-driven actions (such as the configuration of the protection entities and functions, etc.). The presence of a distributed control plane in an MPLS-TP network is optional. However, the absence of such a control plane does not affect the operation of the network and the use of MPLS-TP forwarding, Operations, Administration, and Maintenance (OAM), and survivability capabilities. In particular, the concepts

discussed in [RFC4427] and [RFC4428] refer to recovery actions effected in the data plane; they are equally applicable in MPLS-TP, with or without the use of a control plane.

Thus, some of the MPLS-TP recovery mechanisms do not depend on a control plane and use MPLS-TP OAM mechanisms or management actions to trigger recovery actions.

The principles of MPLS-TP protection-switching actions are similar to those described in [RFC4427], since the protection mechanism is based on the capability to detect certain defects in the transport entities within the recovery domain. The protection-switching controller does not care which initiation method is used, provided that it can be given information about the status of the transport entities within the recovery domain (e.g., OK, signal failure, signal degradation, etc.).

In the context of MPLS-TP, it is imperative to ensure that performing switchovers is possible, regardless of the way in which the network is configured and managed (for example, regardless of whether a control-plane, management-plane, or OAM initiation mechanism is used).

All MPLS and GMPLS protection mechanisms [RFC4428] are applicable in an MPLS-TP environment. It is also possible to provision and manage the related protection entities and functions defined in MPLS and GMPLS using the management plane [RFC5654]. Regardless of whether an OAM, management, or control plane initiation mechanism is used, the protection-switching operation is a data-plane operation.

In some recovery schemes (such as bidirectional protection switching), it is necessary to coordinate the protection state between the edges of the recovery domain to achieve initiation of recovery actions for both directions. An MPLS-TP protocol may be used as an in-band (i.e., data-plane based) control protocol in order to coordinate the protection state between the edges of the protection domain. When the MPLS-TP control plane is in use, a control-plane-based mechanism can also be used to coordinate the protection states between the edges of the protection domain.

1.3. Recovery Context

An MPLS-TP Label Switched Path (LSP) may be subject to any part of or all of MPLS-TP link recovery, path-segment recovery, or end-to-end recovery, where:

- MPLS-TP link recovery refers to the recovery of an individual link (and hence all or a subset of the LSPs routed over the link) between two MPLS-TP nodes. For example, link recovery may be provided by server-layer recovery.

- Segment recovery refers to the recovery of an LSP segment (i.e., segment and concatenated segment in the language of [RFC5654]) between two nodes and is used to recover from the failure of one or more links or nodes.

- End-to-end recovery refers to the recovery of an entire LSP, from its ingress to its egress node.

For additional resiliency, more than one of these recovery techniques may be configured concurrently for a single path.

Co-routed bidirectional MPLS-TP LSPs are defined in a way that allows both directions of the LSP to follow the same route through the network. In this scenario, the operator often requires the directions to fate-share (that is, if one direction fails, both directions should cease to operate).

Associated bidirectional MPLS-TP LSPs exist where the two directions of a bidirectional LSP follow different paths through the network. An operator may also request fate-sharing for associated bidirectional LSPs.

The requirement for fate-sharing causes a direct interaction between the recovery processes affecting the two directions of an LSP, so that both directions of the bidirectional LSP are recovered at the same time. This mode of recovery is termed bidirectional recovery and may be seen as a consequence of fate-sharing.

The recovery scheme operating at the data-plane level can function in a multi-domain environment (in the wider sense of a "domain" [RFC4726]). It can also protect against a failure of a boundary node in the case of inter-domain operation. MPLS-TP recovery schemes are intended to protect client services when they are sent across the MPLS-TP network.

1.4. Scope of This Framework

This framework introduces the architecture of the MPLS-TP recovery domain and describes the recovery schemes in MPLS-TP (based on the recovery types defined in [RFC4427]) as well as the principles of operation, recovery states, recovery triggers, and information exchanges between the different elements that support the reference model.

The framework also describes the qualitative grades of the survivability functions that can be provided, such as dedicated recovery, shared protection, restoration, etc. In the event of a network failure, the grade of recovery directly affects the service grade provided to the end-user.

The general description of the functional architecture is applicable to both LSPs and pseudowires (PWs); however, PW recovery is only introduced in Section 7, and the relevant details are beyond the scope of this document and are for further study.

This framework applies to general recovery schemes as well as to mechanisms that are optimized for specific topologies and are tailored to efficiently handle protection switching.

This document addresses the need for the coordination of protection switching across multiple layers and at sub-layers (for clarity, we use the term "layer" to refer equally to layers and sub-layers). This allows an operator to prevent race conditions and allows the protection-switching mechanism of one layer to recover from a failure before switching is invoked at another layer.

This framework also specifies the functions that must be supported by MPLS-TP to provide the recovery mechanisms. MPLS-TP introduces a tool kit to enable recovery in MPLS-TP-based networks and to ensure that affected services are recovered in the event of a failure.

Generally, network operators aim to provide the fastest, most stable, and best protection mechanism at a reasonable cost in accordance with customer requirements. The greater the grade of protection required, the greater the number of resources will be consumed. It is therefore expected that network operators will offer a wide spectrum of service grade. MPLS-TP-based recovery offers the flexibility to select a recovery mechanism, define the granularity at which traffic delivery is to be protected, and choose the specific traffic types that are to be protected. With MPLS-TP-based recovery, it should be possible to provide different grades of protection for different traffic classes within the same path based on the service requirements.

2. Terminology and References

The terminology used in this document is consistent with that defined in [RFC4427]. The latter is consistent with [G.808.1].

However, certain protection concepts (such as ring protection) are not discussed in [RFC4427]; for those concepts, the terminology used in this document is drawn from [G.841].

Readers should refer to those documents for normative definitions.

This document supplies brief summaries of a number of terms for reasons of clarity and to assist the reader, but it does not redefine terms.

Note, in particular, the distinction and definitions made in [RFC4427] for the following three terms:

- Protection: re-establishing end-to-end traffic delivery using pre-allocated resources.
- Restoration: re-establishing end-to-end traffic delivery using resources allocated at the time of need; sometimes referred to as "repair" of a service, LSP, or the traffic.
- Recovery: a generic term covering both Protection and Restoration.

Note that the term "survivability" is used in [RFC5654] to cover the functional elements of "protection" and "restoration", which are collectively known as "recovery".

Important background information on survivability can be found in [RFC3386], [RFC3469], [RFC4426], [RFC4427], and [RFC4428].

In this document, the following additional terminology is applied:

- "Fault Management", as defined in [RFC5950].
- The terms "defect" and "failure" are used interchangeably to indicate any defect or failure in the sense that they are defined in [G.806]. The terms also include any signal degradation event as defined in [G.806].
- A "fault" is a fault or fault cause as defined in [G.806].
- "Trigger" indicates any event that may initiate a recovery action. See Section 4.1 for a more detailed discussion of triggers.
- The acronym "OAM" is defined as Operations, Administration, and Maintenance, consistent with [RFC6291].
- A "Transport Entity" is a node, link, transport path segment, concatenated transport path segment, or entire transport path.
- A "Working Entity" is a transport entity that carries traffic during normal network operation.

- A "Protection Entity" is a transport entity that is pre-allocated and used to protect and transport traffic when the working entity fails.

- A "Recovery Entity" is a transport entity that is used to recover and transport traffic when the working entity fails.

- "Survivability Actions" are the steps that may be taken by network nodes to communicate faults and to switch traffic from faulted or degraded paths to other paths. This may include sending messages and establishing new paths.

General terminology for MPLS-TP is found in [RFC5921] and [ROSETTA]. Background information on MPLS-TP requirements can be found in [RFC5654].

3. Requirements for Survivability

MPLS-TP requirements are presented in [RFC5654] and serve as normative references for the definition of all MPLS-TP functionality, including survivability. Survivability is presented in [RFC5654] as playing a critical role in the delivery of reliable services, and the requirements for survivability are set out using the recovery terminology defined in [RFC4427].

4. Functional Architecture

This section presents an overview of the elements relating to the functional architecture for survivability within an MPLS-TP network. The components are presented separately to demonstrate the way in which they may be combined to provide the different grades of recovery needed to meet the requirements set out in the previous section.

4.1. Elements of Control

Recovery is achieved by implementing specific actions. These actions aim to repair network resources or redirect traffic along paths that avoid failures in the network. They may be triggered automatically by the MPLS-TP network nodes upon detection of a network defect, or they may be triggered by an operator. Automated actions may be enhanced by in-band (i.e., data-plane-based) OAM mechanisms, or by in-band or out-of-band control-plane signaling.

4.1.1. Operator Control

The survivability behavior of the network as a whole, and the reaction of each transport path when a fault is reported, may be controlled by the operator. This control can be split into two sets of functions: policies and actions performed when the transport path is set up, and commands used to control or force recovery actions for established transport paths.

The operator may establish network-wide or local policies that determine the actions that will be taken when various defects are reported that affect different transport paths. Also, when a service request is made that causes the establishment of one or more transport paths in the network, the operator (or requesting application) may define a particular grade of service, and this will be mapped to specific survivability actions taken before and during transport path setup, after the discovery of a failure of network resources, and upon recovery of those resources.

It should be noted that it is unusual to present a user or customer with options directly related to recovery actions. Instead, the user/customer enters into an SLA with the network provider, and the network operator maps the terms of the SLA (for example, for guaranteed delivery, availability, or reliability) to recovery schemes within the network.

The operator can also issue commands to control recovery actions and events. For example, the operator may perform the following actions:

- Enable or disable the survivability function.
- Invoke the simulation of a network fault.
- Force a switchover from a working path to a recovery path or vice versa.

Forced switchover may be performed for network optimization purposes with minimal service interruption, such as when modifying protected or unprotected services, when replacing MPLS-TP network nodes, etc. In some circumstances, a fault may be reported to the operator, and the operator may then select and initiate the appropriate recovery action. A description of the different operator commands is found in Section 4.12 of [RFC4427].

4.1.2. Defect-Triggered Actions

Survivability actions may be directly triggered by network defects. This means that the device that detects the defect (for example, notification of an issue reported from equipment in a lower layer, failure to receive an OAM Continuity message, or receipt of an OAM message reporting a failure condition) may immediately perform a survivability action.

The action is directly triggered by events in the data plane. Note, however, that coordination of recovery actions between the edges of the recovery domain may require message exchanges for some recovery functions or for performing a bidirectional recovery action.

4.1.3. OAM Signaling

OAM signaling refers to data-plane OAM message exchange. Such messages may be used to detect and localize faults or to indicate a degradation in the operation of the network. However, in this context these messages are used to control or trigger survivability actions. The mechanisms to achieve this are discussed in [RFC6371].

OAM signaling may also be used to coordinate recovery actions within the protection domain.

4.1.4. Control-Plane Signaling

Control-plane signaling is responsible for setup, maintenance, and teardown of transport paths that do not fall under management-plane control. The control plane may also be used to coordinate the detection, localization, and reaction to network defects pertaining to peer relationships (neighbor-to-neighbor or end-to-end). Thus, control-plane signaling may initiate and coordinate survivability actions.

The control plane can also be used to distribute topology and information relating to resource availability. In this way, the "graceful shutdown" [RFC5817] of resources may be affected by withdrawing them; this can be used to invoke a survivability action in a similar way to that used when reporting or discovering a fault, as described in the previous sections.

The use of a control plane for MPLS-TP is discussed in [RFC6373].

4.2. Recovery Scope

This section describes the elements of recovery. These are the quantitative aspects of recovery, that is, the parts of the network for which recovery can be provided.

Note that the terminology in this section is consistent with [RFC4427]. Where the terms differ from those in [RFC5654], mapping is provided.

4.2.1. Span Recovery

A span is a single hop between neighboring MPLS-TP nodes in the same network layer. A span is sometimes referred to as a link, and this may cause some confusion between the concept of a data link and a traffic engineering (TE) link. LSPs traverse TE links between neighboring MPLS-TP nodes in the MPLS-TP network layer. However, a TE link may be provided by any of the following:

- A single data link.
- A series of data links in a lower layer, established as an LSP and presented to the upper layer as a single TE link.
- A set of parallel data links in the same layer, presented either as a bundle of TE links, or as a collection of data links that together provide a data-link-layer protection scheme.

Thus, span recovery may be provided by any of the following:

- Selecting a different TE link from a bundle.
- Moving the TE link so that it is supported by a different data link between the same pair of neighbors.
- Rerouting the LSP in the lower layer.

Moving the protected LSP to another TE link between the same pair of neighbors is a form of segment recovery and not a form of span recovery. Segment Recovery is described in Section 4.2.2.

4.2.2. Segment Recovery

An LSP segment comprises one or more continuous hops on the path of the LSP. [RFC5654] defines two terms. A "segment" is a single hop along the path of an LSP, while a "concatenated segment" is more than one hop along the path of an LSP. In the context of this document, a segment covers both of these concepts.

A PW segment refers to a Single-Segment PW (SS-PW) or to a single segment of a Multi-Segment PW (MS-PW) that is set up between two PE devices that may be Terminating PEs (T-PEs) or Switching PEs (S-PEs) so that the full set of possibilities is T-PE to S-PE, S-PE to S-PE, S-PE to T-PE, or T-PE to T-PE (for the SS-PW case). As indicated in Section 1, the recovery of PWs and PW segments is beyond the scope of this document; however, see Section 7.

Segment recovery involves redirecting or copying traffic at the source end of a segment onto an alternate path leading to the other end of the segment. According to the required grade of recovery (described in Section 4.3), traffic may be either redirected to a pre-established segment, through rerouting the protected segment, or tunneled to the far end of the protected segment through a "bypass" LSP. For details on recovery mechanisms, see Section 4.4.

Note that protecting a transport path against node failure requires the use of segment recovery or end-to-end recovery, while a link failure can be protected using span, segment, or end-to-end recovery.

4.2.3. End-to-End Recovery

End-to-end recovery is a special case of segment recovery where the protected segment comprises the entire transport path. End-to-end recovery may be provided as link-diverse or node-diverse recovery where the recovery path shares no links or no nodes with the working path.

Note that node-diverse paths are necessarily link-diverse and that full, end-to-end node-diversity is required to guarantee recovery.

Two observations need to be made about end-to-end recovery.

- Firstly, there may be circumstances where node-diverse end-to-end paths do not guarantee recovery. The ingress and egress nodes will themselves be single points of failure. Additionally, there may be shared risks of failure (for example, geographic collocation, shared resources, etc.) between diverse nodes as described in Section 4.9.2.

- Secondly, it is possible to use end-to-end recovery techniques even when there is not full diversity and the working and protection paths share links or nodes.

4.3. Grades of Recovery

This section describes the qualitative grades of survivability that can be provided. In the event of a network failure, the grade of recovery offered directly affects the service grade provided to the end-user. This will be observed as the amount of data lost when a network fault occurs, and the length of time required to recover connectivity.

In general, there is a correlation between the recovery service grade (i.e., the speed of recovery and reduction of data loss) and the amount of resources used in the network; better service grades require the pre-allocation of resources to the recovery paths, and those resources cannot be used for other purposes if high-quality recovery is required. An operator will consider how providing different grades of recovery may require that network resources be provisioned and allocated for exclusive use of the recovery paths such that the resources cannot be used to support other customer services.

Sections 6 and 7 of [RFC4427] provide a full breakdown of the protection and recovery schemes. This section summarizes the qualitative grades available.

Note that, in the context of recovery, a useful discussion of the term "resource" and its interpretation in both the IETF and ITU-T contexts may be found in Section 3.2 of [RFC4397].

The selection of the recovery grade and schemes to satisfy the service grades for an LSP using available network resources is subject to network and local policy and may be pre-designated through network planning or may be dynamically determined by the network.

4.3.1. Dedicated Protection

In dedicated protection, the resources for the recovery entity are pre-assigned for the sole use of the protected transport path. This will clearly be the case in 1+1 protection, and may also be the case in 1:1 protection where extra traffic (see Section 4.3.3) is not supported.

Note that when using protection tunnels (see Section 4.4.3), resources may also be dedicated to the protection of a specific transport path. In some cases (1:1 protection), the entire bypass tunnel may be dedicated to providing recovery for a specific transport path, while in other cases (such as facility backup), a subset of the resources associated with the bypass tunnel may be pre-assigned for the recovery of a specific service.

However, as described in Section 4.4.3, the bypass tunnel method can also be used for shared protection (Section 4.3.2), either to carry extra traffic (Section 4.3.3) or to achieve best-effort recovery without the need for resource reservation.

4.3.2. Shared Protection

In shared protection, the resources for the recovery entities of several services are shared. These may be shared as 1:n or m:n and are shared on individual links. Link-by-link resource sharing may be managed and operated along LSP segments, on PW segments, or on end-to-end transport paths (LSP or PW). Note that there is no requirement for m:n recovery in the list of MPLS-TP requirements documented in [RFC5654]. Shared protection can be applied in different topologies (mesh, ring, etc.) and can utilize different protection mechanisms (linear, ring, etc.).

End-to-end shared protection shares resources between a number of paths that have common end points. Thus, a number of paths (n paths) are all protected by one or more protection paths (m paths, where m may equal 1). When there have been m failures, there are no more available protection paths, and the n paths are no longer protected. Thus, in 1:n protection, one fault can be protected against before all the n paths are unprotected. The fact that the paths have become unprotected needs to be conveyed to the path end points since they may need to report the change in service grade or may need to take further action to increase their protection. In end-to-end shared protection, this communication is simple since the end points are common.

In shared mesh protection (see Section 4.7.6), the paths that share the protection resources do not necessarily have the same end points. This provides a more flexible resource-sharing scheme, but the network planning and the coordination of protection state after a recovery action are more complex.

Where a bypass tunnel is used (Section 4.4.3), the tunnel might not have sufficient resources to simultaneously protect all of the paths for which it offers protection; in the event that all paths were affected by network defects and failures at the same time, not all of them would be recovered. Policy would dictate how this situation should be handled: some paths might be protected, while others would simply fail; the traffic for some paths would be guaranteed, while traffic on other paths would be treated as best-effort with the risk of dropped packets. Alternatively, it is possible that protection would not be attempted according to local policy at the nodes that perform the recovery actions.

Shared protection is a trade-off between assigning network resources to protection (which is not required most of the time) and risking unrecoverable services in the event that multiple network defects or failures occur. Rapid recovery can be achieved with dedicated protection, but it is delayed by message exchanges in the management, control, or data planes for shared protection. This means that there is also a trade-off between rapid recovery and resource sharing. In some cases, shared protection might not meet the speed required for protection, but it may still be faster than restoration.

These trade-offs may be somewhat mitigated by the following:

- Adjusting the value of n in 1:n protection.
- Using m:n protection for a value of m > 1.
- Establishing new protection paths as each available protection path is put into use.

In an MPLS-TP network, the degree to which a resource is shared between LSPs is a policy issue. This policy may be applied to the resource or to the LSPs, and may be pre-configured, configured per LSP and installed during LSP establishment, or may be dynamically configured.

4.3.3. Extra Traffic

Section 2.5.1.1 of [RFC5654] says: "Support for extra traffic (as defined in [RFC4427]) is not required in MPLS-TP and MAY be omitted from the MPLS-TP specifications". This document observes that extra traffic facilities may therefore be provided as part of the MPLS-TP survivability toolkit depending upon the development of suitable solution specifications. The remainder of this section explains the concepts of extra traffic without prejudging the decision to specify or not specify such solutions.

Network resources allocated for protection represent idle capacity during the time that recovery is not actually required, and can be utilized by carrying other traffic, referred to as "extra traffic".

Note that extra traffic does not need to start or terminate at the ends of the entity (e.g., LSP) that it uses.

When a network resource carrying extra traffic is required for the recovery of protected traffic from the failed working path, the extra traffic is disrupted. This disruption make take one of two forms:

- In "hard preemption", the extra traffic is excluded from the protection resource. The disruption of the extra traffic is total, and the service supported by the extra traffic must be dropped, or some form of rerouting or restoration must be applied to the extra traffic LSP in order to recover the service.

  Hard preemption is achieved by "setting a switch" on the path of the extra traffic such that it no longer flows. This situation may be detected by OAM and reported as a fault, or may be proactively reported through OAM or control-plane signaling.

- In "soft preemption", the extra traffic is not explicitly excluded from the protection resource, but is given lower priority than the protected traffic. In a packet network (such as MPLS-TP), this can result in oversubscription of the protection resource with the result that the extra traffic receives "best-effort" delivery. Depending on the volume of protection and extra traffic, and the level of oversubscription, the extra traffic may be slightly or heavily impacted.

  The event of soft preemption may be detected by OAM and reported as a degradation of traffic delivery or as a fault. It may also be proactively reported through OAM or control-plane signaling.

Note that both hard and soft preemption may utilize additional message exchanges in the management, control, or data planes. These messages do not necessarily mean that recovery is delayed, but may increase the complexity of the protection system. Thus, the benefits of carrying extra traffic must be weighed against the disadvantages of delayed recovery, additional network overhead, and the impact on the services that support the extra traffic according to the details of the solutions selected.

Note that extra traffic is not protected by definition, but may be restored.

Extra traffic is not supported on dedicated protection resources, which, by definition, are used for 1+1 protection (Section 4.3.1), but it can be supported in other protection schemes, including shared protection (Section 4.3.2) and tunnel protection (Section 4.4.3).

Best-effort traffic should not be confused with extra traffic. For best-effort traffic, the network does not guarantee data delivery, and the user does not receive guaranteed quality of service (e.g., in terms of jitter, packet loss, delay, etc.). Best-effort traffic depends on the current traffic load. However, for extra traffic, quality can only be guaranteed until resources are required for recovery. At this point, the extra traffic may be completely

displaced, may be treated as best effort, or may itself be recovered (for example, by restoration techniques).

4.3.4. Restoration

This section refers to LSP restoration. Restoration for PWs is beyond the scope of this document (but see Section 7).

Restoration represents the most effective use of network resources, since no resources are reserved for recovery. However, restoration requires the computation of a new path and the activation of a new LSP (through the management or control plane). It may be more time-consuming to perform these steps than to implement recovery using protection techniques.

Furthermore, there is no guarantee that restoration will be able to recover the service. It may be that all suitable network resources are already in use for other LSPs, so that no new path can be found. This problem can be partially mitigated by using LSP setup priorities, so that recovery LSPs can preempt existing LSPs with lower priorities.

Additionally, when a network defect occurs, multiple LSPs may be disrupted by the same event. These LSPs may have been established by different Network Management Stations (NMSes) or they may have been signaled by different head-end MPLS-TP nodes, meaning that multiple points in the network will try to compute and establish recovery LSPs at the same time. This can lead to a lack of resources within the network and cause recovery failures; some recovery actions will need to be retried, resulting in even slower recovery times for some services.

Both hard and soft LSP restoration may be supported. For hard LSP restoration, the resources of the working LSP are released before the recovery LSP is fully established (i.e., break-before-make). For soft LSP restoration, the resources of the working LSP are released after an alternate LSP is fully established (i.e., make-before-break). Note that in the case of reversion (Section 4.3.5), the resources associated with the working LSP are not released.

The restoration resources may be pre-calculated and even pre-signaled before the restoration action starts, but not pre-allocated. This is known as pre-planned LSP restoration. The complete establishment/activation of the restoration LSP occurs only when the restoration action starts. Pre-planning may occur periodically and provides the most accurate information about the available resources in the network.

4.3.5. Reversion

After a service has been recovered and traffic is flowing along the recovery LSP, the defective network resource may be replaced. Traffic can be redirected back onto the original working LSP (known as "reversion"), or it can be left where it is on the recovery LSP ("non-revertive" behavior).

It should be possible to specify the reversion behavior of each service; this might even be configured for each recovery instance.

In non-revertive mode, an additional operational option is possible where protection roles are switched, so that the recovery LSP becomes the working LSP, while the previous working path (or the resources used by the previous working path) are used for recovery in the event of an additional fault.

In revertive mode, it is important to prevent excessive swapping between the working and recovery paths in the case of an intermittent defect. This can be addressed by using a reversion delay timer (the Wait-To-Restore timer), which controls the length of time to wait before reversion following the repair of a fault on the original working path. It should be possible for an operator to configure this timer per LSP, and a default value should be defined.

4.4. Mechanisms for Protection

This section provides general descriptions (MPLS-TP non-specific) of the mechanisms that can be used for protection purposes. As indicated above, while the functional architecture applies to both LSPs and PWs, the mechanism for recovery described in this document refers to LSPs and LSP segments only. Recovery mechanisms for pseudowires and pseudowire segments are for further study and will be described in a separate document (see also Section 7).

4.4.1. Link-Level Protection

Link-level protection refers to two paradigms: (1) where protection is provided in a lower network layer and (2) where protection is provided by the MPLS-TP link layer.

Note that link-level protection mechanisms do not protect the nodes at each end of the entity (e.g., a link or span) that is protected. End-to-end or segment protection should be used in conjunction with link-level protection to protect against a failure of the edge nodes.

Link-level protection offers the following grades of protection:

- Full protection where a dedicated protection entity (e.g., a link or span) is pre-established to protect a working entity. When the working entity fails, the protected traffic is switched to the protecting entity. In this scenario, all LSPs carried over the working entity are recovered (in one protection operation) when there is a failure condition. This is referred to in [RFC4427] as "bulk recovery".

- Partial protection where only a subset of the LSPs or traffic carried over a selected entity is recovered when there is a failure condition. The decision as to which LSPs will be recovered and which will not depends on local policy.

When there is no failure on the working entity, the protection entity may transport extra traffic that may be preempted when protection switching occurs.

If link-level protection is available, it may be desirable to allow this to be attempted before attempting other recovery mechanisms for the transport paths affected by the fault because link-level protection may be faster and more conservative of network resources. This can be achieved both by limiting the propagation of fault condition notifications and by delaying the other recovery actions. This consideration of other protection can be compared with the discussion of recovery domains (Section 4.5) and recovery in multi-layer networks (Section 4.9).

A protection mechanism may be provided at the MPLS-TP link layer (which connects two MPLS-TP nodes). Such a mechanism can make use of the procedures defined in [RFC5586] to set up in-band communication channels at the MPLS-TP Section level, to use these channels to monitor the health of the MPLS-TP link, and to coordinate the protection states between the ends of the MPLS-TP link.

4.4.2. Alternate Paths and Segments

The use of alternate paths and segments refers to the paradigm whereby protection is performed in the network layer in which the protected LSP is located; this applies either to the entire end-to-end LSP or to a segment of the LSP. In this case, hierarchical LSPs are not used (compare with Section 4.4.3).

Different grades of protection may be provided:

- Dedicated protection where a dedicated entity (e.g., LSP or LSP segment) is (fully) pre-established to protect a working entity

(e.g., LSP or LSP segment). When a failure condition occurs on the working entity, traffic is switched onto the protection entity. Dedicated protection may be performed using 1:1 or 1+1 linear protection schemes. When the failure condition is eliminated, the traffic may revert to the working entity. This is subject to local configuration.

- Shared protection where one or more protection entities is pre-established to protect against a failure of one or more working entities (1:n or m:n).

When the fault condition on the working entity is eliminated, the traffic should revert back to the working entity in order to allow other related working entities to be protected by the shared protection resource.

4.4.3. Protection Tunnels

A protection tunnel is pre-provisioned in order to protect against a failure condition along a sequence of spans in the network. This may be achieved using LSP heirarchy. We call such a sequence a network segment. A failure of a network segment may affect one or more LSPs that transit the network segment.

When a failure condition occurs in the network segment (detected either by OAM on the network segment, or by OAM on a concatenated segment of one of the LSPs transiting the network segment), one or more of the protected LSPs are switched over at the ingress point of the network segment and are transmitted over the protection tunnel. This is implemented through label stacking. Label mapping may be an option as well.

Different grades of protection may be provided:

- Dedicated protection where the protection tunnel reserves sufficient resources to provide protection for all protected LSPs without causing service degradation.

- Partial protection where the protection tunnel has enough resources to protect some of the protected LSPs, but not all of them simultaneously. Policy dictates how this situation should be handled: it is possible that some LSPs would be protected, while others would simply fail; it is possible that traffic would be guaranteed for some LSPs, while for other LSPs it would be treated as best effort with the risk of packets being dropped. Alternatively, it is possible that protection would not be attempted.

4.5. Recovery Domains

Protection and restoration are performed in the context of a recovery domain. A recovery domain is defined between two or more recovery reference end points that are located at the edges of the recovery domain and that border on the element on which recovery can be provided (as described in Section 4.2). This element can be an end-to-end path, a segment, or a span.

An end-to-end path can be observed as a special segment case where the ingress and egress Label Edge Routers (LERs) serve as the recovery reference end points.

In this simple case of a point-to-point (P2P) protected entity, two end points reside at the boundary of the protection domain. An LSP can enter through one reference end point and exit the recovery domain through another reference end point.

In the case of unidirectional point-to-multipoint (P2MP), three or more end points reside at the boundary of the protection domain. One of the end points is referred to as the source/root, while the others are referred to as sinks/leaves. An LSP can enter the recovery domain through the root point and exit the recovery domain through the leaf points.

The recovery mechanism should restore traffic that was interrupted by a facility (link or node) fault within the recovery domain. Note that a single link may be part of several recovery domains. If two recovery domains have common links, one recovery domain must be contained within the other. This can be referred to as nested recovery domains. The boundaries of recovery domains may coincide, but recovery domains must not overlap.

Note that the edges of a recovery domain are not protected, and unless the whole domain is contained within another recovery domain, the edges form a single point of failure.

A recovery group is defined within a recovery domain and consists of a working (primary) entity and one or more recovery (backup) entities that reside between the end points of the recovery domain. To guarantee protection in all situations, a dedicated recovery entity should be pre-provisioned using disjoint resources in the recovery domain, in order to protect against a failure of a working entity. Of course, mechanisms to detect faults and to trigger protection switching are also needed.

The method used to monitor the health of the recovery element is beyond the scope of this document. The end points that are

responsible for the recovery action must receive information on its condition. The condition of the recovery element may be 'OK', 'failed', or 'degraded'.

When the recovery operation is to be triggered by OAM mechanisms, an OAM Maintenance Entity Group must be defined for each of the working and protection entities.

The recovery entities and functions in a recovery domain can be configured using a management plane or a control plane. A management plane may be used to configure the recovery domain by setting the reference points, the working and recovery entities, and the recovery type (e.g., 1:1 bidirectional linear protection, ring protection, etc.). Additional parameters associated with the recovery process may also be configured. For more details, see Section 6.1.

When a control plane is used, the ingress LERs may communicate with the recovery reference points that request that protection or restoration be configured across a recovery domain. For details, see Section 6.5.

Cases of multiple interconnections between distinct recovery domains create a hierarchical arrangement of recovery domains, since a single top-level recovery domain is created from the concatenation of two recovery domains with multiple interconnections. In this case, recovery actions may be taken both in the individual, lower-level recovery domains to protect any LSP segment that crosses the domain, and within the higher-level recovery domain to protect the longer LSP segment that traverses the higher-level domain.

The MPLS-TP recovery mechanism can be arranged to ensure coordination between domains. In interconnected rings, for example, it may be preferable to allow the upstream ring to perform recovery before the downstream ring, in order to ensure that recovery takes place in the ring in which the defect occurred. Coordination of recovery actions is particularly important in nested domains and is discussed further in Section 4.9.

4.6. Protection in Different Topologies

As described in the requirements listed in Section 3 and detailed in [RFC5654], the selected recovery techniques may be optimized for different network topologies if the optimized mechanisms perform significantly better than the generic mechanisms in the same topology.

These mechanisms are required (R91 of [RFC5654]) to interoperate with the mechanisms defined for arbitrary topologies, in order to allow

end-to-end protection and to ensure that consistent protection techniques are used across the entire network. In this context, 'interoperate' means that the use of one technique must not inhibit the use of another technique in an adjacent part of the network for use on the same end-to-end transport path, and must not prohibit the use of end-to-end protection mechanisms.

The next sections (4.7 and 4.8) describe two different topologies and explain how recovery may be markedly different in those different scenarios. They also develop the concept of a recovery domain and show how end-to-end survivability may be achieved through a concatenation of recovery domains, each providing some grade of recovery in part of the network.

4.7. Mesh Networks

A mesh network is any network where there is arbitrary interconnectivity between nodes in the network. Mesh networks are usually contrasted with more specific topologies such as hub-and-spoke or ring (see Section 4.8), although such networks are actually examples of mesh networks. This section is limited to the discussion of protection techniques in the context of mesh networks. That is, it does not include optimizations for specific topologies.

Linear protection is a protection mechanism that provides rapid and simple protection switching. In a mesh network, linear protection provides a very suitable protection mechanism because it can operate between any pair of points within the network. It can protect against a defect in a node, a span, a transport path segment, or an end-to-end transport path. Linear protection gives a clear indication of the protection status.

Linear protection operates in the context of a protection domain. A protection domain is a special type of recovery domain (see Section 4.5) associated with the protection function. A protection domain is composed of the following architectural elements:

- A set of end points that reside at the boundary of the protection domain. In the simple case of 1:n or 1+1 P2P protection, two end points reside at the boundary of the protection domain. In each transmission direction, one of the end points is referred to as the source, and the other is referred to as the sink. For unidirectional P2MP protection, three or more end points reside at the boundary of the protection domain. One of the end points is referred to as the source/root, while the others are referred to as sinks/leaves.

- o A Protection Group consists of one or more working (primary) paths and one or more protection (backup) paths that run between the end points belonging to the protection domain. To guarantee protection in all scenarios, a dedicated protection path should be pre-provisioned to protect against a defect of a working path (i.e., 1:1 or 1+1 protection schemes). In addition, the working and the protection paths should be disjoint; i.e., the physical routes of the working and the protection paths should be physically diverse in every respect.

Note that if the resources of the protection path are less than those of the working path, the protection path may not have sufficient resources to protect the traffic of the working path.

As mentioned in Section 4.3.2, the resources of the protection path may be shared as 1:n. In this scenario, the protection path will not have sufficient resources to protect all the working paths at a specific time.

For bidirectional P2P paths, both unidirectional and bidirectional protection switching are supported. If a defect occurs when bidirectional protection switching is defined, the protection actions are performed in both directions (even if the defect is unidirectional). The protection state is required to operate with a level of coordination between the end points of the protection domain.

In unidirectional protection switching, the protection actions are only performed in the affected direction.

Revertive and non-revertive operations are provided as options for the network operator.

Linear protection supports the protection schemes described in the following sub-sections.

4.7.1. 1:n Linear Protection

In the 1:1 scheme, a protection path is allocated to protect against a defect, failure, or a degradation in a working path. As described above, to guarantee protection, the protection entity should support the full capacity and bandwidth, although it may be configured (for example, because of limited network resource availability) to offer a degraded service when compared with the working entity.

Figure 1 presents 1:1 protection architecture. In normal conditions, data traffic is transmitted over the working entity, while the protection entity functions in the idle state. (OAM may run on the

protection entity to verify its state.) Normal conditions are defined when there is no defect, failure, or degradation on the working entity, and no administrative configuration or request causes traffic to flow over the protection entity.

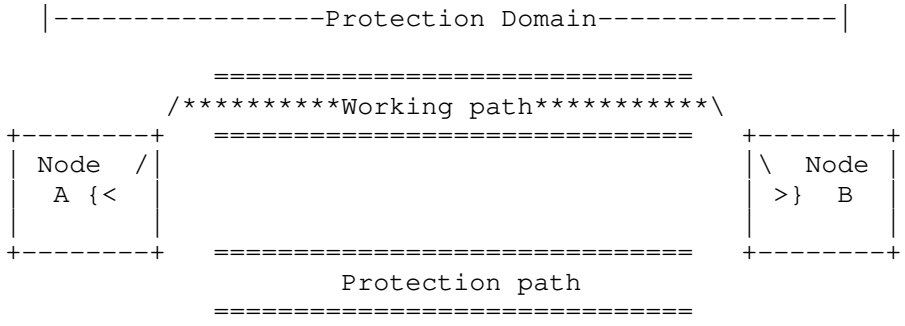

Figure 1: 1:1 Protection Architecture

If there is a defect on the working entity or a specific administrative request, traffic is switched to the protection entity.

Note that when operating with non-revertive behavior (see Section 4.3.5), after the conditions causing the switchover have been cleared, the traffic continues to flow on the protection path, but the working and protection roles are not switched.

In each transmission direction, the protection domain source bridges traffic onto the appropriate entity, while the sink selects traffic from the appropriate entity. The source and the sink need to coordinate the protection states to ensure that bridging and selection are performed to and from the same entity. For this reason, a signaling coordination protocol (either a data-plane in-band signaling protocol or a control-plane-based signaling protocol) is required.

In bidirectional protection switching, both ends of the protection domain are switched to the protection entity (even when the fault is unidirectional). This requires a protocol to coordinate the protection state between the two end points of the protection domain.

When there is no defect, the bandwidth resources of the idle entity may be used for traffic with lower priority. When protection switching is performed, the traffic with lower priority may be preempted by the protected traffic through tearing down the LSP with lower priority, reporting a fault on the LSP with lower priority, or by treating the traffic with lower priority as best effort and discarding it when there is congestion.

In the general case of 1:n linear protection, one protection entity is allocated to protect n working entities. The protection entity might not have sufficient resources to protect all the working entities that may be affected by fault conditions at a specific time. In this case, in order to guaranteed protection, the protection entity should support enough capacity and bandwidth to protect any of the n working entities.

When defects or failures occur along multiple working entities, the entity to be protected should be prioritized. The protection states between the edges of the protection domain should be fully coordinated to ensure consistent behavior. As explained in Section 4.3.5, revertive behavior is recommended when 1:n is supported.

4.7.2. 1+1 Linear Protection

In the 1+1 protection scheme, a fully dedicated protection entity is allocated.

As depicted in Figure 2, data traffic is copied and fed at the source to both the working and the protection entities. The traffic on the working and the protection entities is transmitted simultaneously to the sink of the protection domain, where selection between the working and protection entities is performed (based on some predetermined criteria).

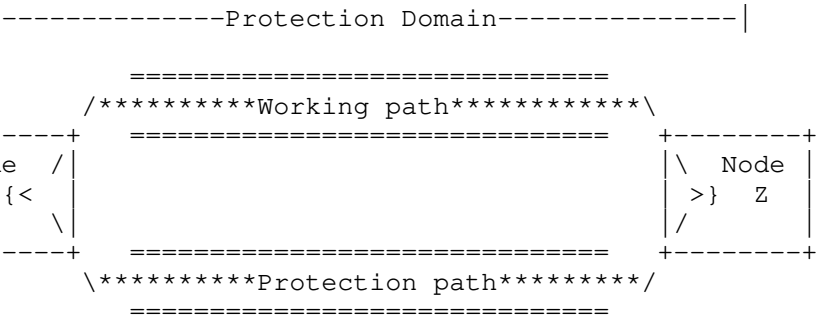

Figure 2: 1+1 Protection Architecture

Note that control traffic between the edges of the protection domain (such as OAM or a control protocol to coordinate the protection state, etc.) may be transmitted on an entity that differs from the one used for the protected traffic. These packets should not be discarded by the sink.

In 1+1 unidirectional protection switching, there is no need to coordinate the protection state between the protection controllers at both ends of the protection domain. In 1+1 bidirectional protection switching, a protocol is required to coordinate the protection state between the edges of the protection domain.

In both protection schemes, traffic flows end-to-end on the working entity after the conditions causing the switchover have been cleared. Data selection may return to selecting traffic from the working entity if reversion is enabled, and it will require coordination of the protection state between the edges of the protection domain. To avoid frequent switching caused by intermittent defects or failures when the network is not stable, traffic is not selected from the working entity before the Wait-To-Restore (WTR) timer has expired.

4.7.3. P2MP Linear Protection

Linear protection may be applied to protect unidirectional P2MP entities using 1+1 protection architecture. The source/root MPLS-TP node bridges the user traffic to both the working and protection entities. Each sink/leaf MPLS-TP node selects the traffic from one entity according to some predetermined criteria. Note that when there is a fault condition on one of the branches of the P2MP path, some leaf MPLS-TP nodes may select the working entity, while other leaf MPLS-TP nodes may select traffic from the protection entity.

In a 1:1 P2MP protection scheme, the source/root MPLS-TP node needs to identify the existence of a fault condition on any of the branches of the network. This means that the sink/leaf MPLS-TP nodes need to notify the source/root MPLS-TP node of any fault condition. This also necessitates a return path from the sinks/leaves to the source/root MPLS-TP node. When protection switching is triggered, the source/root MPLS-TP node selects the protection transport path for traffic transfer.

A form of "segment recovery for P2MP LSPs" could be constructed. Given a P2MP LSP, one can protect any possible point of failure (link or node) using N backup P2MP LSPs. Each backup P2MP LSP originates from the upstream node with respect to a different possible failure point and terminates at all of the destinations downstream of the potential failure point. In case of a failure, traffic is redirected to the backup P2MP path.

Note that such mechanisms do not yet exist, and their exact behavior is for further study.

A 1:n protection scheme for P2MP transport paths is also required by [RFC5654]. Such a mechanism is for future study.

4.7.4. Triggers for the Linear Protection Switching Action

Protection switching may be performed when:

- A defect condition is detected on the working entity, and the protection entity has "no" or an inferior condition. Proactive in-band OAM Continuity Check and Connectivity Verification (CC-V) monitoring of both the working and the protection entities may be used to enable the rapid detection of a fault condition. For protection switching, it is common to run a CC-V every 3.33 ms. In the absence of three consecutive CC-V messages, a fault condition is declared. In order to monitor the working and the protection entities, an OAM Maintenance Entity Group should be defined for each entity. OAM indications associated with fault conditions should be provided at the edges of the protection domain that is responsible for the protection-switching operation. Input from OAM performance monitoring that indicates degradation in the working entity may also be used as a trigger for protection switching. In the case of degradation, switching to the protection entity is needed only if the protection entity can exhibit better operating conditions.

- An indication is received from a lower-layer server that there is a defect in the lower layer.

- An external operator command is received (e.g., 'Forced Switch', 'Manual Switch'). For details, see Section 6.1.2.

- A request to switch over is received from the far end. The far end may initiate this request, for example, on receipt of an administrative request to switch over, or when bidirectional 1:1 protection switching is supported and a defect occurred that could only be detected by the far end, etc.

As described above, the protection state should be coordinated between the end points of the protection domain. Control messages should be exchanged between the edges of the protection domain to coordinate the protection state of the edge nodes. Control messages can be delivered using an in-band, data-plane-driven control protocol or a control-plane-based protocol.

For 50-ms protection switching, it is recommended that an in-band, data-plane-driven signaling protocol be used in order to coordinate the protection states. An in-band, data-plane protocol for use in MPLS-TP networks is documented in [MPLS-TP-LP] for linear protection (ring protection is discussed in Section 4.8 of this document). This protocol is also used to detect mismatches between the configurations provisioned at the ends of the protection domain.

As described in Section 6.5, the GMPLS control plane already includes procedures and message elements to coordinate the protection states between the edges of the protection domain. These procedures and protocol messages are specified in [RFC4426], [RFC4872], and [RFC4873]. However, these messages lack the capability to coordinate the revertive/non-revertive behavior and the consistency of configured timers at the edges of the protection domain (timers such as WTR, hold-off timer, etc.).

4.7.5. Applicability of Linear Protection for LSP Segments

In order to implement data-plane-based linear protection on LSP segments, use is made of the Sub-Path Maintenance Element (SPME), an MPLS-TP architectural element defined in [RFC5921]. Maintenance operations (e.g., monitoring, protection, or management) engage with message transmission (e.g., OAM, Protection Path Coordination, etc.) in the maintained domain. Further discussion of the architecture for OAM and SPME is found in [RFC5921] and [RFC6371]. An SPME is an LSP that is basically defined and used for the purposes of OAM monitoring, protection, or management of LSP segments. The SPME uses the MPLS construct of a hierarchical, nested LSP, as defined in [RFC3031].

For linear protection, SPMEs should be defined over the working and protection entities between the edges of a protection domain. OAM messages and messages used to coordinate protection state can be initiated at the edge of the SPME and sent to the peer edge of the SPME. Note that these messages are sent over the Generic Associated Channel (G-ACh) within the SPME, and that they use a two-label stack, the SPME label, and, at the bottom of the stack, the G-ACh label (GAL) [RFC5586].

The end-to-end traffic of the LSP, which includes data traffic and control traffic (messages for OAM, management, signaling, and to coordinate protection state), is tunneled within the SPMEs by means of label stacking, as defined in [RFC3031].

Mapping between an LSP and an SPME can be 1:1; this is similar to the ITU-T Tandem Connection element that defines a sub-layer corresponding to a segment of a path. Mapping can also be 1:n to allow the scalable protection of a set of LSP segments traversing the part of the network in which a protection domain is defined. Note that each of these LSPs can be initiated or terminated at different end points in the network, but that they all traverse the protection domain and share similar constraints (such as requirements for quality of service (QoS), terms of protection, etc.).

Note also that in the context of segment protection, the SPMEs serve as the working and protection entities.

4.7.6. Shared Mesh Protection

For shared mesh protection, the protection resources are used to protect multiple LSPs that do not all share the same end points; for example, in Figure 3 there are two paths, ABCDE and VWXYZ. These paths do not share end points and cannot, therefore, make use of 1:n linear protection, even though they do not have any common points of failure.

ABCDE may be protected by the path APQRE, while VWXYZ can be protected by the path VPQRZ. In both cases, 1:1 or 1+1 protection may be used. However, it can be seen that if 1:1 protection is used for both paths, the PQR network segment does not carry traffic when no failures affect either of the two working paths. Furthermore, in the event of only one failure, the PQR segment carries traffic from only one of the working paths.

Thus, it is possible for the network resources on the PQR segment to be shared by the two recovery paths. In this way, mesh protection can substantially reduce the number of network resources that have to be reserved in order to provide 1:n protection.

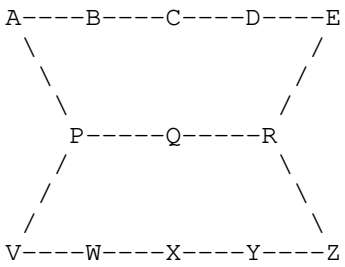

Figure 3: A Shared Mesh Protection Topology

As the network becomes more complex and the number of LSPs increases, the potential for shared mesh protection also increases. However, this can quickly become unmanageable owing to the increased complexity. Therefore, shared mesh protection is normally pre-planned and configured by the operator, although an automated system cannot be ruled out.

Note that shared mesh protection operates as 1:n linear protection (see Section 4.7.1). However, the protection state needs to be coordinated between a larger number of nodes: the end points of the shared concatenated protection segment (nodes P and R in the example)

as well as the end points of the protected LSPs (nodes A, E, V, and Z in the example).

Additionally, note that the shared-protection resources could be used to carry extra traffic. For example, in Figure 4, an LSP JPQRK could be a preemptable LSP that constitutes extra traffic over the PQR hops; it would be displaced in the event of a protection event. In this case, it should be noted that the protection state must also be coordinated with the ends of the extra-traffic LSPs.

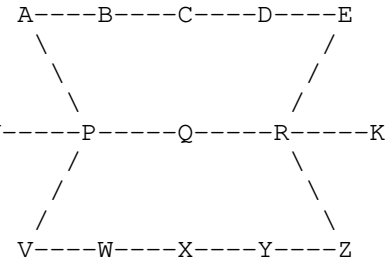

Figure 4: Shared Mesh Protection with Extra Traffic

4.8. Ring Networks

Several service providers have expressed great interest in the operation of MPLS-TP in ring topologies; they demand a high degree of survivability functionality in these topologies.

Various criteria for optimization are considered in ring topologies, such as:

1. Simplification in ring operation in terms of the number of OAM Maintenance Entities that are needed to trigger the recovery actions, the number of recovery elements, the number of management-plane transactions during maintenance operations, etc.

2. Optimization of resource consumption around the ring, such as the number of labels needed for the protection paths that traverse the network, the total bandwidth required in the ring to ensure path protection, etc. (see R91 of [RFC5654]).

[RFC5654] introduces a list of requirements for ring protection covering the recovery mechanisms needed to protect traffic in a single ring as well as traffic that traverses more than one ring. Note that configuration and the operation of the recovery mechanisms in a ring must scale well with the number of transport paths, the number of nodes, and the number of ring interconnects.

The requirements for ring protection are fully compatible with the generic requirements for recovery.

The architecture and the mechanisms for ring protection are specified in separate documents. These mechanisms need to be evaluated against the requirements specified in [RFC5654], which includes guidance on the principles for the development of new mechanisms.

4.9. Recovery in Layered Networks

In multi-layer or multi-regional networking [RFC5212], recovery may be performed at multiple layers or across nested recovery domains.

The MPLS-TP recovery mechanism must ensure that the timing of recovery is coordinated in order to avoid race scenarios. This also allows the recovery mechanism of the server layer to fix the problem before recovery takes place in the MPLS-TP layer, or the MPLS-TP layer to perform recovery before a client network.

A hold-off timer is required to coordinate recovery timing in multiple layers or across nested recovery domains. Setting this configurable timer involves a trade-off between rapid recovery and the creation of a race condition where multiple layers respond to the same fault, potentially allocating resources in an inefficient manner. Thus, the detection of a defect condition in the MPLS-TP layer should not immediately trigger the recovery process if the hold-off timer is configured as a value other than zero. Instead, the hold-off timer should be started when the defect is detected and, on expiry, the recovery element should be checked to determine whether the defect condition still exists. If it does exist, the defect triggers the recovery operation.

The hold-off timer should be configurable.

In other configurations, where the lower layer does not have a restoration capability, or where it is not expected to provide protection, the lower layer needs to trigger the higher layer to immediately perform recovery. Although this can be forced by configuring the hold-off timer as zero, it may be that because of layer independence, the higher layer does not know whether the lower layer will perform restoration. In this case, the higher layer will configure a non-zero hold-off timer and rely on the receipt of a specific notification from the lower layer if the lower layer cannot perform restoration. Since layer boundaries are always within nodes, such coordination is implementation-specific and does not need to be covered here.

Reference should be made to [RFC3386], which discusses the interaction between layers in survivable networks.

4.9.1. Inherited Link-Level Protection

Where a link in the MPLS-TP network is formed through connectivity (i.e., a packet or non-packet LSP) in a lower-layer network, that connectivity may itself be protected; for example, the LSP in the lower-layer network may be provisioned with 1+1 protection. In this case, the link in the MPLS-TP network has an inherited grade of protection.

An LSP in the MPLS-TP network may be provisioned with protection in the MPLS-TP network, as already described, or it may be provisioned to utilize only those links that have inherited protection.

By classifying the links in the MPLS-TP network according to the grade of protection that they inherited from the server network, it is possible to compute an end-to-end path in the MPLS-TP network that uses only those links with a specific or superior grade of inherited protection. This means that the end-to-end MPLS-TP LSP can be protected at the grade necessary to conform to the SLA without needing to provide any additional protection in the MPLS-TP layer. This reduces complexity, saves network resources, and eliminates protection-switching coordination problems.

When the requisite grade of inherited protection is not available on all segments along the path in the MPLS-TP network, segment protection may be used to achieve the desired protection grade.

It should be noted, however, that inherited protection only applies to links. Nodes cannot be protected in this way. An operator will need to perform an analysis of the relative likelihood and consequences of node failure if this approach is taken without providing protection in the MPLS-TP LSP or PW layer to handle node failure.

4.9.2. Shared Risk Groups

When an MPLS-TP protection scheme is established, it is important that the working and protection paths do not share resources in the network. If this is not achieved, a single defect may affect both the working and the protection paths with the result that traffic cannot be delivered -- since under such a condition the traffic was not protected.

Note that this restriction does not apply to restoration, since this takes place after the fault has occurred, which means that the point of failure can be avoided if an available path exists.

When planning a recovery scheme, it is possible to use a topology map of the MPLS-TP layer to select paths that use diverse links and nodes within the MPLS-TP network. However, this does not guarantee that the paths are truly diverse; for example, two separate links in an MPLS-TP network may be provided by two lambdas in the same optical fiber, or by two fibers that cross the same bridge. Moreover, two completely separate MPLS-TP nodes might be situated in the same building with a shared power supply.

Thus, in order to achieve proper recovery planning, the MPLS-TP network must have an understanding of the groups of lower-layer resources that share a common risk of failure. From this, MPLS-TP shared risk groups can be constructed that show which MPLS-TP resources share a common risk of failure. Diversity of working and protection paths can be planned, not only with regard to nodes and links but also in order to refrain from using resources from the same shared risk groups.

4.9.3. Fault Correlation

In a layered network, a low-layer fault may be detected and reported by multiple layers and may sometimes lead to the generation of multiple fault reports from the same layer. For example, a failure of a data link may be reported by the line cards in an MPLS-TP node, but it could also be detected and reported by the MPLS-TP OAM.

Section 4.6 explains how it is important to coordinate the survivability actions configured and operated in a multi-layer network in a way that will avoid over-equipping the survivability resources in the network, while ensuring that recovery actions are performed in only one layer at a time.

Fault correlation is about understanding which single event has generated a set of fault reports, so that recovery actions can be coordinated, and so that the fault logging system does not become overloaded. Fault correlation depends on understanding resource use at lower layers, shared risk groups, and a wider view with regard to the way in which the layers are interrelated.

Fault correlation is most easily performed at the point of fault detection; for example, an MPLS-TP node that receives a fault notification from the lower layer, and detects a fault on an LSP in the MPLS-TP layer, can easily correlate these two events. Furthermore, if the same node detects multiple faults on LSPs that

share the same faulty data link, it can easily correlate them. Such a node may use correlation to perform group-based recovery actions and can reduce the number of alarm events that it generates to its management station.

Fault correlation may also be performed at a management station that receives fault reports from different layers and different nodes in the network. This enables the management station to coordinate management-originated recovery actions and to present consolidated fault information to the user and automated management systems.

It is also necessary to correlate fault information detected and reported through OAM. This function would enable a fault detected at a lower layer, and reported at a transit node of an MPLS-TP LSP, to be correlated with an MPLS-TP-layer fault detected at a Maintenance End Point (MEP) -- for example, the egress of the MPLS-TP LSP. Such correlation allows the coordination of recovery actions performed at the MEP, but it also requires that the lower-layer fault information is propagated to the MEP, which is most easily achieved using a control plane, management plane, or OAM message.

5. Applicability and Scope of Survivability in MPLS-TP

The MPLS-TP network can be viewed as two layers (the MPLS LSP layer and the PW layer). The MPLS-TP network operates over data-link connections and data-link networks whereby the MPLS-TP links are provided by individual data links or by connections in a lower-layer network. The MPLS LSP layer is a mandatory part of the MPLS-TP network, while the PW layer is an optional addition for supporting specific services.

MPLS-TP survivability provides recovery from failure of the links and nodes in the MPLS-TP network. The link defects and failures are typically caused by defects or failures in the underlying data-link connections and networks, but this section is only concerned with recovery actions performed in the MPLS-TP network, which must recover from the manifestation of any problem as a defect failure in the MPLS-TP network.

This section lists the recovery elements (see Section 1) supported in each of the two layers that can recover from defects or failures of nodes or links in the MPLS-TP network.

Recovery Element	MPLS LSP Layer	PW Layer
Link Recovery	MPLS LSP recovery can be used to survive the failure of an MPLS-TP link.	The PW layer is not aware of the underlying network. This function is not supported.
Segment/Span Recovery	An individual LSP segment can be recovered to survive the failure of an MPLS-TP link.	For an SS-PW, segment recovery is the same as end-to-end recovery. Segment recovery for an MS-PW is for future study, and this function is now provided using end-to-end recovery.
Concatenated Segment Recovery	A concatenated LSP segment can be recovered to survive the failure of an MPLS-TP link or node.	Concatenated segment recovery (in an MS-PW) is for future study, and this function is now provided using end-to-end recovery.
End-to-End Recovery	An end-to-end LSP can be recovered to survive any node or link failure, except for the failure of the ingress or egress node.	End-to-end PW recovery can be applied to survive any node (including S-PE) or link failure, except for failure of the ingress or egress T-PE.
Service Recovery	The MPLS LSP layer is service-agnostic. This function is not supported.	PW-layer service recovery requires surviving faults in T-PEs or on Attachment Circuits (ACs). This is currently out of scope for MPLS-TP.

Table 1: Recovery Elements Supported
by the MPLS LSP Layer and PW Layer

Section 6 provides a description of mechanisms for MPLS-TP-LSP survivability. Section 7 provides a brief overview of mechanisms for MPLS-TP-PW survivability.

6. Mechanisms for Providing Survivability for MPLS-TP LSPs

This section describes the existing mechanisms that provide LSP protection within MPLS-TP networks and highlights areas where new work is required.

6.1. Management Plane

As described above, a fundamental requirement of MPLS-TP is that recovery mechanisms should be capable of functioning in the absence of a control plane. Recovery may be triggered by MPLS-TP OAM fault management functions or by external requests (e.g., an operator's request for manual control of protection switching). Recovery LSPs (and in particular Restoration LSPs) may be provisioned through the management plane.

The management plane may be used to configure the recovery domain by setting the reference end-point points (which control the recovery actions), the working and the recovery entities, and the recovery type (e.g., 1:1 bidirectional linear protection, ring protection, etc.).

Additional parameters associated with the recovery process (such as WTR and hold-off timers, revertive/non-revertive operation, etc.) may also be configured.

In addition, the management plane may initiate manual control of the recovery function. A priority should be set for the fault conditions and the operator's requests.

Since provisioning the recovery domain involves the selection of a number of options, mismatches may occur at the different reference points. The MPLS-TP protocol to coordinate protection state, which is specified in [MPLS-TP-LP], may be used as an in-band (i.e., data-plane-based) control protocol to coordinate the protection states between the end points of the recovery domain, and to check the consistency of configured parameters (such as timers, revertive/non-revertive behavior, etc.) with discovered inconsistencies that are reported to the operator.

It should also be possible for the management plane to track the recovery status by receiving reports or by issuing polls.

6.1.1. Configuration of Protection Operation

To implement the protection-switching mechanisms, the following entities and information should be configured and provisioned:

- The end points of a recovery domain. As described above, these end points border on the element of recovery to which recovery is applied.

- The protection group, which, depending on the required protection scheme, consists of a recovery entity and one or more working entities. In 1:1 or 1+1 P2P protection, the paths of the working entity and the recovery entities must be physically diverse in every respect (i.e., not share any resources or physical locations), in order to guarantee protection.

- As defined in Section 4.8, the SPME must be supported in order to implement data-plane-based LSP segment recovery, since related control messages (e.g., for OAM, Protection Path Coordination, etc.) can be initiated and terminated at the edges of a path where push and pop operations are enabled. The SPME is an end-to-end LSP that in this context corresponds to the recovery entities (working and protection) and makes use of the MPLS construct of hierarchical nested LSP, as defined in [RFC3031]. OAM messages and messages to coordinate protection state can be initiated at the edge of the SPME and sent over G-ACH to the peer edge of the SPME. It is necessary to configure the related SPMEs and map between the LSP segments being protected and the SPME. Mapping can be 1:1 or 1:N to allow scalable protection of a set of LSP segments traversing the part of the network in which a protection domain is defined.

  Note that each of these LSPs can be initiated or terminated at different end points in the network, but that they all traverse the protection domain and share similar constraints (such as requirements for QoS, terms of protection, etc.).

- The protection type that should be defined (e.g., unidirectional 1:1, bidirectional 1+1, etc.)

- Revertive/non-revertive behavior should be configured.

- Timers (such as WTR, hold-off timer, etc.) should be set.

6.1.2. External Manual Commands

The following external, manual commands may be provided for manual control of the protection-switching operation. These commands apply to a protection group; they are listed in descending order of priority:

- Blocked protection action - a manual command to prevent data traffic from switching to the recovery entity. This command actually disables the protection group.

- Force protection action - a manual command that forces a switch of normal data traffic to the recovery entity.

- Manual protection action - a manual command that forces a switch of data traffic to the recovery entity only when there is no defect in the recovery entity.

- Clear switching command - the operator may request that a previous administrative switch command (manual or force switch) be cleared.

6.2. Fault Detection

Fault detection is a fundamental part of recovery and survivability. In all schemes, with the exception of some types of 1+1 protection, the actions required for the recovery of traffic delivery depend on the discovery of some kind of fault. In 1+1 protection, the selector (at the receiving end) may simply be configured to choose the better signal; thus, it does not detect a fault or degradation of itself, but simply identifies the path that is better for data delivery.

Faults may be detected in a number of ways depending on the traffic pattern and the underlying hardware. End-to-end faults may be reported by the application or by knowledge of the application's data pattern, but this is an unusual approach. There are two more common mechanisms for detecting faults in the MPLS-TP layer:

- Faults reported by the lower layers.

- Faults detected by protocols within the MPLS-TP layer.

In an IP/MPLS network, the second mechanism may utilize control-plane protocols (such as the routing protocols) to detect a failure of adjacency between neighboring nodes. In an MPLS-TP network, it is possible that no control plane will be present. Even if a control plane is present, it will be a GMPLS control plane [RFC3945], which logically separates control channels from data channels; thus, no conclusion about the health of a data channel can be drawn from the

failure of an associated control channel. MPLS-TP-layer faults are, therefore, only detected through the use of OAM protocols, as described in Section 6.4.1.

Faults may, however, be reported by a lower layer. These generally show up as interface failures or data-link failures (sometimes known as connectivity failures) within the MPLS-TP network, for example, an underlying optical link may detect loss of light and report a failure of the MPLS-TP link that uses it. Alternatively, an interface card failure may be reported to the MPLS-TP layer.

Faults reported by lower layers are only visible in specific nodes within the MPLS-TP network (i.e., at the adjacent end points of the MPLS-TP link). This would only allow recovery to be performed locally, so, to enable recovery to be performed by nodes that are not immediately local to the fault, the fault must be reported (Sections 6.4.3 and 6.5.4).

6.3. Fault Localization

If an MPLS-TP node detects that there is a fault in an LSP (that is, not a network fault reported from a lower layer, but a fault detected by examining the LSP), it can immediately perform a recovery action. However, unless the location of the fault is known, the only practical options are:

- Perform end-to-end recovery.

- Perform some other recovery as a speculative act.

Since the speculative acts are not guaranteed to achieve the desired results and could consume resources unnecessarily, and since end-to-end recovery can require a lot of network resources, it is important to be able to localize the fault.

Fault localization may be achieved by dividing the network into protection domains. End-to-end protection is thereby operated on LSP segments, depending on the domain in which the fault is discovered. This necessitates monitoring of the LSP at the domain edges.

Alternatively, a proactive mechanism of fault localization through OAM (Section 6.4.3) or through the control plane (Section 6.5.3) is required.

Fault localization is particularly important for restoration because a new path must be selected that avoids the fault. It may not be practical or desirable to select a path that avoids the entire failed

working path, and it is therefore necessary to isolate the fault's location.

6.4. OAM Signaling

MPLS-TP provides a comprehensive set of OAM tools for fault management and performance monitoring at different nested levels (end-to-end, a portion of a path (LSP or PW), and at the link level) [RFC6371].

These tools support proactive and on-demand fault management (for fault detection and fault localization) as well as performance monitoring (to measure the quality of the signals and detect degradation).

To support fast recovery, it is useful to use some of the proactive tools to detect fault conditions (e.g., link/node failure or degradation) and to trigger the recovery action.

The MPLS-TP OAM messages run in-band with the traffic and support unidirectional and bidirectional P2P paths as well as P2MP paths.

As described in [RFC6371], MPLS-TP OAM operates in the context of a Maintenance Entity that borders on the OAM responsibilities and represents the portion of a path between two points that is monitored and maintained, and along which OAM messages are exchanged. [RFC6371] refers also to a Maintenance Entity Group (MEG), which is a collection of one or more Maintenance Entities (MEs) that belong to the same transport path (e.g., P2MP transport path) and which are maintained and monitored as a group.

An ME includes two MEPs (Maintenance Entity Group End Points) that reside at the boundaries of an ME, and a set of zero or more MIPs (Maintenance Entity Group Intermediate Points) that reside within the Maintenance Entity along the path. A MEP is capable of initiating and terminating OAM messages, and as such can only be located at the edges of a path where push and pop operations are supported. In order to define an ME over a portion of path, it is necessary to support SPMEs.

The SPME is an end-to-end LSP that in this context corresponds to the ME; it uses the MPLS construct of hierarchical nested LSPs, which is defined in [RFC3031]. OAM messages can be initiated at the edge of the SPME and sent over G-ACH to the peer edge of the SPME.

The related SPMEs must be configured, and mapping must be performed between the LSP segments being monitored and the SPME. Mapping can be 1:1 or 1:N to allow scalable operation. Note that each of these

LSPs can be initiated or terminated at different end points in the network and can share similar constraints (such as requirements for QoS, terms of protection, etc.).

With regard to recovery, where MPLS-TP OAM is supported, an OAM Maintenance Entity Group is defined for each of the working and protection entities.

6.4.1. Fault Detection

MPLS-TP OAM tools may be used proactively to detect the following fault conditions between MEPs:

- Loss of continuity and misconnectivity - the proactive Continuity Check (CC) function is used to detect loss of continuity between two MEPs in an MEG. The proactive Connectivity Verification (CV) allows a sink MEP to detect a misconnectivity defect (e.g., mismerge or misconnection) with its peer source MEP when the received packet carries an incorrect ME identifier. For protection switching, it is common to run a CC-V (Continuity Check and Connectivity Verification) message every 3.33 ms. In the absence of three consecutive CC-V messages, loss of continuity is declared and is notified locally to the edge of the recovery domain in order to trigger a recovery action. In some cases, when a slower recovery time is acceptable, it is also possible to lengthen the transmission rate.

- Signal degradation - notification from OAM performance monitoring indicating degradation in the working entity may also be used as a trigger for protection switching. In the event of degradation, switching to the recovery entity is necessary only if the recovery entity can guarantee better conditions. Degradation can be measured by proactively activating MPLS-TP OAM packet loss measurement or delay measurement.

- A MEP can receive an indication from its sink MEP of a Remote Defect Indication and locally notify the end point of the recovery domain regarding the fault condition, in order to trigger the recovery action.

6.4.2. Testing for Faults

The management plane may be used to initiate the testing of links, LSP segments, or entire LSPs.

MPLS-TP provides OAM tools that may be manually invoked on-demand for a limited period, in order to troubleshoot links, LSP segments, or entire LSPs (e.g., diagnostics, connectivity verification, packet

loss measurements, etc.). On-demand monitoring covers a combination of "in-service" and "out-of-service" monitoring functions. Out-of-service testing is supported by the OAM on-demand lock operation. The lock operation temporarily disables the transport entity (LSP, LSP segment, or link), preventing the transmission of all types of traffic, with the exceptions of test traffic and OAM (dedicated to the locked entity).

[RFC6371] describes the operations of the OAM functions that may be initiated on-demand and provides some considerations.

MPLS-TP also supports in-service and out-of-service testing of the recovery (protection and restoration) mechanism, the integrity of the protection/recovery transport paths, and the coordination protocol between the end points of the recovery domain. The testing operation emulates a protection-switching request but does not perform the actual switching action.

6.4.3. Fault Localization

MPLS-TP provides OAM tools to locate a fault and determine its precise location. Fault detection often only takes place at key points in the network (such as at LSP end points or at MEPs). This means that a fault may be located anywhere within a segment of the relevant LSP. Finer information granularity is needed to implement optimal recovery actions or to diagnose the fault. On-demand tools like trace-route, loopback, and on-demand CC-V can be used to localize a fault.

The information may be notified locally to the end point of the recovery domain to allow implementation of optimal recovery action. This may be useful for the re-calculation of a recovery path.

The information should also be reported to network management for diagnostic purposes.

6.4.4. Fault Reporting

The end points of a recovery domain should be able to detect fault conditions in the recovery domain and to notify the management plane.

In addition, a node within a recovery domain that detects a fault condition should also be able to report this to network management. Network management should be capable of correlating the fault reports and identifying the source of the fault.

MPLS-TP OAM tools support a function where an intermediate node along a path is able to send an alarm report message to the MEP, indicating

the presence of a fault condition in the server layer that connects it to its adjacent node. This capability allows a MEP to suppress alarms that may be generated as a result of a failure condition in the server layer.

6.4.5. Coordination of Recovery Actions

As described above, in some cases (such as in bidirectional protection switching, etc.) it is necessary to coordinate the protection states between the edges of the recovery domain. [MPLS-TP-LP] defines procedures, protocol messages, and elements for this purpose.

The protocol is also used to signal administrative requests (e.g., manual switch, etc.), but only when these are provisioned at the edge of the recovery domain.

The protocol also enables mismatches to be detected between the configurations at the ends of the protection domain (such as timers, revertive/non-revertive behavior); these mismatches can subsequently be reported to the management plane.

In the absence of suitable coordination (owing to failures in the delivery or processing of the coordination protocol messages), protection switching will fail. This means that the operation of the protocol that coordinates the protection state is a fundamental part of protection switching.

6.5. Control Plane

The GMPLS control plane has been proposed as the control plane for MPLS-TP [RFC5317]. Since GMPLS was designed for use in transport networks, and since it has been implemented and deployed in many networks, it is not surprising that it contains many features that support a high degree of survivability.

The signaling elements of the GMPLS control plane utilize extensions to the Resource Reservation Protocol (RSVP) (as described in a series of documents commencing with [RFC3471] and [RFC3473]), although it is based on [RFC3209] and [RFC2205]. The architecture for GMPLS is provided in [RFC3945], while [RFC4426] gives a functional description of the protocol extensions needed to support GMPLS-based recovery (i.e., protection and restoration).

A further control-plane protocol called the Link Management Protocol (LMP) [RFC4204] is part of the GMPLS protocol family and can be used to coordinate fault localization and reporting.

Clearly, the control-plane techniques described here only apply where an MPLS-TP control plane is deployed and operated. All mandatory MPLS-TP survivability features must be enabled, even in the absence of the control plane. However, when present, the control plane may be used to provide alternative mechanisms that may be desirable, since they offer simple automation or a richer feature set.

6.5.1. Fault Detection

The control plane is unable to detect data-plane faults. However, it does provide mechanisms that detect control-plane faults, and these can be used to recognize data-plane faults when it is evident that the control and data planes are fate-sharing. Although [RFC5654] specifies that MPLS-TP must support an out-of-band control channel, it does not insist that it be used exclusively. This means that there may be deployments where an in-band (or at least an in-fiber) control channel is used. In this scenario, failure of the control channel can be used to infer that there is a failure of the data channel, or, at least, it can be used to trigger an investigation of the health of the data channel.

Both RSVP and LMP provide a control channel "keep-alive" mechanism (called the Hello message in both cases). Failure to receive a message in the configured/negotiated time period indicates a control-plane failure. GMPLS routing protocols ([RFC4203] and [RFC5307]) also include keep-alive mechanisms designed to detect routing adjacency failures. Although these keep-alive mechanisms tend to operate at a relatively low frequency (on the order of seconds), it is still possible that the first indication of a control-plane fault will be received through the routing protocol.

Note, however, that care must be taken to ascertain that a specific failure is not caused by a problem in the control-plane software or in a processor component at the far end of a link.

Because of the various issues involved, it is not recommended that the control plane be used as the primary mechanism for fault detection in an MPLS-TP network.

6.5.2. Testing for Faults

The control plane may be used to initiate and coordinate the testing of links, LSP segments, or entire LSPs. This is important in some technologies where it is necessary to halt data transmission while testing, but it may also be useful where testing needs to be specifically enabled or configured.

LMP provides a control-plane mechanism to test the continuity and connectivity (and naming) of individual links. A single management operation is required to initiate the test at one end of the link, while the LMP handles the coordination with the other end of the link. The test mechanism for an MPLS packet link relies on the LMP Test message inserted into the data stream at one end of the link and extracted at the other end of the link. This mechanism need not disrupt data flowing over the link.

Note that a link in the LMP may, in fact, be an LSP tunnel used to form a link in the MPLS-TP network.

GMPLS signaling (RSVP) offers two mechanisms that may also assist with fault testing. The first mechanism [RFC3473] defines the Admin_Status object that allows an LSP to be set into "testing mode". The interpretation of this mode is implementation-specific and could be documented more precisely for MPLS-TP. The mode sets the whole LSP into a state where it can be tested; this need not be disruptive to data traffic.

The second mechanism provided by GMPLS to support testing is described in [GMPLS-OAM]. This protocol extension supports the configuration (including enabling and disabling) of OAM mechanisms for a specific LSP.

6.5.3. Fault Localization

Fault localization is the process whereby the exact location of a fault is determined. Fault detection often only takes place at key points in the network (such as at LSP end points or at MEPs). This means that a fault may be located anywhere within a segment of the relevant LSP.

If segment or end-to-end protection is in use, this level of information is often sufficient to repair the LSP. However, if finer information granularity is required (either to implement optimal recovery actions or to diagnose a fault), it is necessary to localize the specific fault.

LMP provides a cascaded test-and-propagate mechanism that is designed specifically for this purpose.

6.5.4. Fault Status Reporting

GMPLS signaling uses the Notify message to report fault status [RFC3473]. The Notify message can apply to a single LSP or can carry fault information for a set of LSPs, in order to improve the scalability of fault notification.

Since the Notify message is targeted at a specific node, it can be delivered rapidly without requiring hop-by-hop processing. It can be targeted at LSP end points or at segment end points (such as MEPs). The target points for Notify messages can be manually configured within the network, or they may be signaled when the LSP is set up.

This enables the process to be made consistent with segment protection as well as with the concept of Maintenance Entities.

GMPLS signaling also provides a slower, hop-by-hop mechanism for reporting individual LSP faults on a hop-by-hop basis using PathErr and ResvErr messages.

[RFC4783] provides a mechanism to coordinate alarms and other event or fault information through GMPLS signaling. This mechanism is useful for understanding the status of the resources used by an LSP and for providing information as to why an LSP is not functioning; however, it is not intended to replace other fault-reporting mechanisms.

GMPLS routing protocols [RFC4203] and [RFC5307] are used to advertise link availability and capabilities within a GMPLS-enabled network. Thus, the routing protocols can also provide indirect information about network faults; that is, the protocol may stop advertising or may withdraw the advertisement for a failed link, or it may advertise that the link is about to be shut down gracefully [RFC5817]. This mechanisms is, however, not normally considered to be fast enough for use as a trigger for protection switching.

6.5.5. Coordination of Recovery Actions

Fault coordination is an important feature for certain protection mechanisms (such as bidirectional 1:1 protection). The use of the GMPLS Notify message for this purpose is described in [RFC4426]; however, specific message field values have not yet been defined for this operation.

Further work is needed in GMPLS for control and configuration of reversion behavior for end-to-end and segment protection, and the coordination of timer values.

6.5.6. Establishment of Protection and Restoration LSPs

The management plane may be used to set up protection and recovery LSPs, but, when present, the control plane may be used.

Several protocol extensions exist that simplify this process:

- [RFC4872] provides features that support end-to-end protection switching.
- [RFC4873] describes the establishment of a single, segment-protected LSP. Note that end-to-end protection is a special case of segment protection, and [RFC4872] can also be used to provide end-to-end protection.
- [RFC4874] allows an LSP to be signaled with a request that its path exclude specified resources such as links, nodes, and shared risk link groups (SRLGs). This allows a disjoint protection path to be requested or a recovery path to be set up to avoid failed resources.
- Lastly, it should be noted that [RFC5298] provides an overview of the GMPLS techniques available to achieve protection in multi-domain environments.

7. Pseudowire Recovery Considerations

Pseudowires provide end-to-end connectivity over the MPLS-TP network and may comprise a single pseudowire segment, or multiple segments "stitched" together to provide end-to-end connectivity.

The pseudowire may, itself, require protection, in order to meet the service-level guarantees of its SLA. This protection could be provided by the MPLS-TP LSPs that support the pseudowire, or could be a feature of the pseudowire layer itself.

As indicated above, the functional architecture described in this document applies to both LSPs and pseudowires. However, the recovery mechanisms for pseudowires are for further study and will be defined in a separate document by the PWE3 working group.

7.1. Utilization of Underlying MPLS-TP Recovery

MPLS-TP PWs are carried across the network inside MPLS-TP LSPs. Therefore, an obvious way to provide protection for a PW is to protect the LSP that carries it. Such protection can take any of the forms described in this document. The choice of recovery scheme will depend on the required speed of recovery and the traffic loss that is acceptable for the SLA that the PW is providing.

If the PW is a Multi-Segment PW, then LSP recovery can only protect the PW in individual segments. This means that a single LSP recovery action cannot protect against a failure of a PW switching point (an

S-PE), nor can it protect more than one segment at a time, since the LSP tunnel is terminated at each S-PE. In this respect, LSP protection of a PW is very similar to link-level protection offered to the MPLS-TP LSP layer by an underlying network layer (see Section 4.9).

7.2. Recovery in the Pseudowire Layer

Recovery in the PW layer can be provided by simply running separate PWs end-to-end. Other recovery mechanisms in the PW layer, such as segment or concatenated segment recovery, or service-level recovery involving survivability of T-PE or AC faults will be described in a separate document.

As with any recovery mechanism, it is important to coordinate between layers. This coordination is necessary to ensure that actions associated with recovery mechanisms are only performed in one layer at a time (that is, the recovery of an underlying LSP needs to be coordinated with the recovery of the PW itself). It also makes sure that the working and protection PWs do not both use the same MPLS resources within the network (for example, by running over the same LSP tunnel; see also Section 4.9).

8. Manageability Considerations

Manageability of MPLS-TP networks and their functions is discussed in [RFC5950]. OAM features are discussed in [RFC6371].

Survivability has some key interactions with management, as described in this document. In particular:

- Recovery domains may be configured in a way that prevents one-to-one correspondence between the MPLS-TP network and the recovery domains.

- Survivability policies may be configured per network, per recovery domain, or per LSP.

- Configuration of OAM may involve the selection of MEPs; enabling OAM on network segments, spans, and links; and the operation of OAM on LSPs, concatenated LSP segments, and LSP segments.

- Manual commands may be used to control recovery functions, including forcing recovery and locking recovery actions.

See also the considerations regarding security for management and OAM in Section 9 of this document.

9. Security Considerations

This framework does not introduce any new security considerations; general issues relating to MPLS security can be found in [RFC5920].

However, several points about MPLS-TP survivability should be noted here.

- If an attacker is able to force a protection switch-over, this may result in a small perturbation to user traffic and could result in extra traffic being preempted or displaced from the protection resources. In the case of 1:n protection or shared mesh protection, this may result in other traffic becoming unprotected. Therefore, it is important that OAM protocols for detecting or notifying faults use adequate security to prevent them from being used (through the insertion of bogus messages or through the capture of legitimate messages) to falsely trigger a recovery event.

- If manual commands are modified, captured, or simulated (including replay), it might be possible for an attacker to perform forced recovery actions or to impose lock-out. These actions could impact the capability to provide the recovery function and could also affect the normal operation of the network for other traffic. Therefore, management protocols used to perform manual commands must allow the operator to use appropriate security mechanisms. This includes verification that the user who performs the commands has appropriate authorization.

- If the control plane is used to configure or operate recovery mechanisms, the control-plane protocols must also be capable of providing adequate security.

10. Acknowledgments

Thanks to the following people for useful comments and discussions: Italo Busi, David McWalter, Lou Berger, Yaacov Weingarten, Stewart Bryant, Dan Frost, Lievren Levrau, Xuehui Dai, Liu Guoman, Xiao Min, Daniele Ceccarelli, Scott Bradner, Francesco Fondelli, Curtis Villamizar, Maarten Vissers, and Greg Mirsky.

The Editors would like to thank the participants in ITU-T Study Group 15 for their detailed review.

Some figures and text on shared mesh protection were borrowed from [MPLS-TP-MESH] with thanks to Tae-sik Cheung and Jeong-dong Ryoo.

11. References

11.1. Normative References

[G.806] ITU-T, "Characteristics of transport equipment - Description methodology and generic functionality", Recommendation G.806, January 2009.

[G.808.1] ITU-T, "Generic Protection Switching - Linear trail and subnetwork protection", Recommendation G.808.1, December 2003.

[G.841] ITU-T, "Types and Characteristics of SDH Network Protection Architectures", Recommendation G.841, October 1998.

[RFC2205] Braden, R., Ed., Zhang, L., Berson, S., Herzog, S., and S. Jamin, "Resource ReSerVation Protocol (RSVP) -- Version 1 Functional Specification", RFC 2205, September 1997.

[RFC3209] Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V., and G. Swallow, "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC 3209, December 2001.

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

[RFC3473] Berger, L., Ed., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions", RFC 3473, January 2003.

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

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

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

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

[RFC4428] Papadimitriou, D., Ed., and E. Mannie, Ed., "Analysis of Generalized Multi-Protocol Label Switching (GMPLS)-based Recovery Mechanisms (including Protection and Restoration)", RFC 4428, March 2006.

[RFC4873] Berger, L., Bryskin, I., Papadimitriou, D., and A. Farrel, "GMPLS Segment Recovery", RFC 4873, May 2007.

[RFC5307] Kompella, K., Ed., and Y. Rekhter, Ed., "IS-IS Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 5307, October 2008.

[RFC5317] Bryant, S., Ed., and L. Andersson, Ed., "Joint Working Team (JWT) Report on MPLS Architectural Considerations for a Transport Profile", RFC 5317, February 2009.

[RFC5586] Bocci, M., Ed., Vigoureux, M., Ed., and S. Bryant, Ed., "MPLS Generic Associated Channel", RFC 5586, June 2009.

[RFC5654] Niven-Jenkins, B., Ed., Brungard, D., Ed., Betts, M., Ed., Sprecher, N., and S. Ueno, "Requirements of an MPLS Transport Profile", RFC 5654, September 2009.

[RFC5921] Bocci, M., Ed., Bryant, S., Ed., Frost, D., Ed., Levrau, L., and L. Berger, "A Framework for MPLS in Transport Networks", RFC 5921, July 2010.

[RFC5950] Mansfield, S., Ed., Gray, E., Ed., and K. Lam, Ed., "Network Management Framework for MPLS-based Transport Networks", RFC 5950, September 2010.

[RFC6371] Buci, I., Ed. and B. Niven-Jenkins, Ed., "A Framework for MPLS in Transport Networks", RFC 6371, September 2011.

11.2. Informative References

[GMPLS-OAM] Takacs, A., Fedyk, D., and J. He, "GMPLS RSVP-TE extensions for OAM Configuration", Work in Progress, July 2011.

[MPLS-TP-LP] Weingarten, Y., Osborne, E., Sprecher, N., Fulignoli, A., Ed., and Y. Weingarten, Ed., "MPLS-TP Linear Protection", Work in Progress, August 2011.

[MPLS-TP-MESH] Cheung, T. and J. Ryoo, "MPLS-TP Shared Mesh Protection", Work in Progress, April 2011.

[RFC3031] Rosen, E., Viswanathan, A., and R. Callon, "Multiprotocol Label Switching Architecture", RFC 3031, January 2001.

[RFC3386] Lai, W., Ed., and D. McDysan, Ed., "Network Hierarchy and Multilayer Survivability", RFC 3386, November 2002.

[RFC3469] Sharma, V., Ed., and F. Hellstrand, Ed., "Framework for Multi-Protocol Label Switching (MPLS)-based Recovery", RFC 3469, February 2003.

[RFC4397] Bryskin, I. and A. Farrel, "A Lexicography for the Interpretation of Generalized Multiprotocol Label Switching (GMPLS) Terminology within the Context of the ITU-T's Automatically Switched Optical Network (ASON) Architecture", RFC 4397, February 2006.

[RFC4426] Lang, J., Ed., Rajagopalan, B., Ed., and D. Papadimitriou, Ed., "Generalized Multi-Protocol Label Switching (GMPLS) Recovery Functional Specification", RFC 4426, March 2006.

[RFC4726] Farrel, A., Vasseur, J.-P., and A. Ayyangar, "A Framework for Inter-Domain Multiprotocol Label Switching Traffic Engineering", RFC 4726, November 2006.

[RFC4783] Berger, L., Ed., "GMPLS - Communication of Alarm Information", RFC 4783, December 2006.

[RFC4872] Lang, J., Ed., Rekhter, Y., Ed., and D. Papadimitriou, Ed., "RSVP-TE Extensions in Support of End-to-End Generalized Multi-Protocol Label Switching (GMPLS) Recovery", RFC 4872, May 2007.

[RFC4874] Lee, CY., Farrel, A., and S. De Cnodder, "Exclude Routes - Extension to Resource ReserVation Protocol-Traffic Engineering (RSVP-TE)", RFC 4874, April 2007.

[RFC5212] Shiomoto, K., Papadimitriou, D., Le Roux, JL., Vigoureux, M., and D. Brungard, "Requirements for GMPLS-Based Multi-Region and Multi-Layer Networks (MRN/MLN)", RFC 5212, July 2008.

[RFC5298] Takeda, T., Ed., Farrel, A., Ed., Ikejiri, Y., and JP. Vasseur, "Analysis of Inter-Domain Label Switched Path (LSP) Recovery", RFC 5298, August 2008.

[RFC5817] Ali, Z., Vasseur, JP., Zamfir, A., and J. Newton, "Graceful Shutdown in MPLS and Generalized MPLS Traffic Engineering Networks", RFC 5817, April 2010.

[RFC5920] Fang, L., Ed., "Security Framework for MPLS and GMPLS Networks", RFC 5920, July 2010.

[RFC6373] Andersson, L., Ed., Berger, L., Ed., Fang, L., Ed., and Bitar, N., Ed, and E. Gray, Ed., "MPLS-TP Control Plane Framework", RFC 6373, September 2011.

[RFC6291] Andersson, L., van Helvoort, H., Bonica, R., Romascanu, D., and S. Mansfield, "Guidelines for the Use of the "OAM" Acronym in the IETF", BCP 161, RFC 6291, June 2011.

[ROSETTA] Van Helvoort, H., Ed., Andersson, L., Ed., and N. Sprecher, Ed., "A Thesaurus for the Terminology used in Multiprotocol Label Switching Transport Profile (MPLS-TP) drafts/RFCs and ITU-T's Transport Network Recommendations", Work in Progress, June 2011.

Authors' Addresses

Nurit Sprecher (editor)
Nokia Siemens Networks
3 Hanagar St.
Neve Ne'eman B Hod
Hasharon, 45241 Israel

EMail: nurit.sprecher@nsn.com

Adrian Farrel (editor)
Juniper Networks

EMail: adrian@olddog.co.uk