Comments received before and during the Q9 conference call on 16-11-2009 are included as comments and/or editorials throughout this document.

General comments applicable to this draft are:

1. The draft includes a large amount of repetitive material, some restructuring (e.g. introducing the recovery domain earlier, removing detail from the introduction) may help to eliminate this repetition and reduce the overall size of the draft.
2. Here is an inconsistent use of terms, e.g. synchronise ←→ coordinate, span ←→ link, etc.
3. Avoid the use of the term cost in terms of optimisation; it is proposed to use the term complexity.
4. Extra traffic: there is no requirement for support for extra traffic in [RFC4427]. In a ring where MPLS-TP must support the sharing of protection bandwidth thus allowing best-effort traffic (108). In ITU-T this is referred to as NUT (non-preemptible unprotected traffic). It is proposed to remove extra traffic from this draft.
5. The acronyms PSC and APS are used, in ITU-T APS is a generic term used in different methodologies and does not provide a solution.
6. In this draft, there are several sentences that tell this I-D is applicable to PW and LSP. However, PWE3 WG is progressing PW redundancy drafts so it is not clear whether the PW redundancy aspects is covered or not in this Survivability Framework draft. To avoid this, it could be better to indicate that this I-D focuses on only the case of the connection between pair of PEs for recovery. (i.e. the case where AC and CE is included is out of scope od this I-D)

Network Working Group   N. Sprecher
Internet-Draft   Nokia Siemens Networks
Intended status: Informational   A. Adrian Farrel
Expires: April 28, 2010   Old Dog Consulting
                                                          Oct 25, 2009

Multiprotocol Label Switching Transport Profile Survivability Framework
draft-ietf-mpls-tp-survive-fwk-02.txt

Status of this Memo

Copyright Notice

Abstract

Network survivability is the network's ability to restore traffic delivery following failure or degradation of network resources or ann attack on the network. It plays a critical role in the delivery of guaranteed services in transport networks to meet the requirements expressed in Service Level Agreements (SLAs).

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

This document provides a framework for the provision of survivability functions in the data plane of an MPLS-TP network using tools provided by the management plane and/or the control plane as well as autonomous techniques inherent in the data plane itself.

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

Table of Contents

1. Introduction

Network survivability is the network's ability to restore traffic delivery following a failure or degradation of traffic delivery caused by a network fault or an attack on the network; it 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 restore network operations in accordance with the terms of the SLA.

The MPLS Transport Profile (MPLS-TP) is described in [RFC5654] and [MPLS-TP-FWK] . MPLS-TP is designed to be consistent with existing transport network operations and management models, and to provide survivability mechanisms, such as protection and restoration~~. The function provided is intended to be similar or better to that found in established transport networks which set a high benchmark for reliability.~~

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

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 within the agreed service level and to recover from 'failed' or 'degraded' transport entities (links or nodes). Such actions are normally initiated by the detection of a defect or performance degradation, or by an external request (e.g. an operator request for manual control of protection switching).

[RFC4427] makes a distinction between protection switching and restoration mechanisms. Protection switching is an autonomous process and triggered without managemenet or control plane involvement. It makes use of pre-assigned capacity between nodes, where the simplest scheme has one dedicated protection entity for each working entity, while the most complex scheme has m protection entities shared between n working entities (m:n). Protection switching may be either unidirectional or bidirectional; unidirectional meaning that each direction of a bidirectional connection is protection switched independently, while bidirectional means that both directions are switched at the same time even if the fault applies to only one direction of the connection. Restoration uses any capacity available between nodes and usually involves re-routing. The resources used for restoration may be pre-planned and recovery priority may be used as a differentiation mechanism to determine which services are recovered and which are not recovered or are sacrificed in order to achieve

Comment [M1]: High level comment: It may be better to significantly reduce the level of detail in the introduction i.e. address the comments in this section by removing text vs. adding more details.

Comment [G2]: Generaql – no common terminology is used through the whole document. The words fault, failure, protection, recovery and survivability are used interchangeably

Comment [HvH3]: Provide more detail and align with Rosetta draft anomaly/defect/failure

Comment [M4]: The objective is that MPLS-TP is consistent with the operation model of existing transport networks. Therefore "better" implies different and may miss the objective of operational compatibility

Comment [M5]: Need to distinguish between defined – i.e. resources are identified but not used and assigned – resources cannot be used for other purposes (1:1 vs. 1+1).

Comment [M6]: Applies to restoration as well – should also mention revertive

Comment [M7]: How is pre planned different from protection?

recovery of other services. In general, protection actions are completed within time frames of tens of milliseconds, while automated restoration actions are normally completed in periods ranging from hundreds of milliseconds to a maximum of a few seconds.

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, and the absence of such a control plane does not affect the ability to operate the network and to use MPLS-TP forwarding, Operations, Administration and Maintenance (OAM), and survivability capabilities.

Thus, some of the MPLS-TP recovery mechanisms do not depend on a control plane and use of MPLS-TP OAM mechanisms or management actions to trigger protection switching across connections that were set up using management plane configuration. These OAM mechanisms may be triggered by data plane events or by operator actions, and are based on MPLS-TP OAM fault management functions. 'Fault management' in this context refers to failure detection, localization, and notification (where the term 'failure' is used to represent both signal failure and signal degradation). The term 'trigger' is used to indicate any event that may be used to activate ~~cause an implementation to consider taking~~ protection action.

The principles of MPLS-TP protection switching operation are similar to those described in [RFC4427] as the protection mechanism is based on the ability to detect certain defects in the transport entities within the recovery domain. In the context of MPLS-TP, transport entities are nodes, links, concatenated segments, connection and transport paths. The protection switching controller does not care which monitoring method is used, as long as it can be given information about the status of the transport entities within the recovery domain (e.g. 'OK', signal failure, signal degradation, etc.).

The protection switching operation is basically a data-plane capability and in the context of MPLS-TP it needs to be ensured that it is possible to switch over independent of the way the network is configured and managed. All the MPLS and GMPLS protection mechanisms are applicable in MPLS-TP environment, and it should be possible also to provision and manage the related protection entities and functions defined in MPLS and GMPLS using a management plane.

In some protection switching schemes (such as bidirectional protection switching), it is necessary to coordinate the protection state between the edges of the recovery domain. An MPLS-TP Protection State Coordination (PSC) protocol may be used as an in-

Comment [M8]: So which aspects of 4427 & 4428 are applicable to MPLS-TP?
Do we plan to define restoration without a control plane – possible but it would require 10s of seconds if OSS based.

Comment [M9]: Protection vs. restoration?

Comment [M10]: How do operator actions trigger an trigger an OAM mechanism? Are you considering APS signalling as an OAM mechanism?

Comment [M11]: Localozation is only required to the level of a recovery domain – no need to localize to a node/link to trigger protection.

Comment [M12]: As stated above 4427 assumes a control plane.

Comment [M13]: Align with the main frame work and OAM framework drafts.

Comment [M14]: A node is not a transport entity

Comment [M15]: Not if they rely on IP forwarding. May be used optionally.

Comment [M16]: Also applies to restoration – If this level of detail is provided in the introduction it should also cover revertive recovery.

band (i.e. data plane-based) control protocol to align both ends of the protected domain. Control plane-based mechanism can also be used to synchronize the protection states between the edges of the protection domain.

The MPLS-TP recovery mechanisms may be applied at various nested levels throughout the MPLS-TP network, as is the case with the recovery schemes defined in [RFC4427] and [RFC4873]. A Label Switching Path (LSP) may be subject to any or all of MPLS-TP link recovery, path segment recovery, or end-to-end recovery, where:

- o 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.
- o Segment recovery refers to the recovery of an LSP segment (i.e. segment and concatenated segment in the language of [RFC5654]) between two nodes which are the boundary nodes of the segment.
- o End-to-end recovery refers to the recovery of an entire LSP from its ingress to its egress node.

Multiple recovery levels may be used concurrently by a single LSP for added resiliency. Co-routed bidirectional MPLS-TP LSPs are defined such that both directions of the LSP follow the same route through the network. In this case the directions are often required by the operator to fate-share (that is, if one direction fails, both directions should cease to operate). This may also be the case for associated bidirectional LSPs where the two directions of the LSP take different paths through the network. This causes a direct interaction between the recovery levels affecting the directions of an LSP such that both directions of the LSP are switched to a new MPLS-TP link, segment, or end-to-end path together.

The recovery scheme operating at the data plane level can function in a multi-domain environment; it should 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 traffic as it is sent across the MPLS-TP network. This document introduces protection and restoration techniques in general terms and then describes how they may be applied in the LSP layer and in the pseudowire layer to meet the requirements of the MPLS-TP recovery schemes [RFC5654]. A description of the MPLS-TP LSP and pseudowire layers can be found in [MPLS-TP-FWK].

Comment [YT17]: there seems to be the overlapping item between single (one hop) segment and {span or link}
For those clause, it could be better restructure
- link
- single segment
- multiple segments (i.e. TCM)

Comment [M18]: Is this framework intended to apply to PWs and MS-PWs?

Comment [M19]: If the link is recovered (vs. the segments on the link) how can only a subset be recovered?

Comment [M20]: Adjacent nodes, or ?

Comment [M21]: The term recovery domain is introduced much later in the draft – are these nodes at the boundary of the recovery domain?
In which case segment = domain?

Comment [HvH22]: What is meant by level in this context, is it nested protection?

Comment [M23]: If recovery has been successful at one "level" why invoke another level?
Should describe in terms of nested and concatenated domains. For a single fault recovery would only be invoked in one domain. Also need to describe measures to avoid "flap" i.e. multiple domains attempting recovery simultaneously.
(GA: you should avoid multiple recovery)

Comment [M24]: No need to force a "fault" on the direction that has not failed.
GA: This is to switch both paths together protection and working paths rather than cease perhaps free the associated path

Comment [M25]: Since the two directions are routed independently why would we need to invoke bi directional recovery?

Comment [M26]: How? This is a per recovery domain at the same "level"

Comment [M27]: The new segment must also be co-routed if the original path was co-routed.

Comment [M28]: Need to distinguish between the types of domain – e.g. administrative domain and recovery domain. A recovery domain may "enclose" more than one administrative domain and thus provide recovery for failures of nodes or links on the boundary between administrative domains. But by definition recovery is limited to the scope/extent of the recovery domain.
(GA: does multi-domain involves different transport technology. It should be clarified)

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 sustain the reference model. The reference model is based on the MPLS-TP OAM reference model which is defined in [MPLS-TP-OAM].

The framework also describes the qualitative levels of the survivability functions that can be provided, such as dedicated recovery, shared protection, restoration, etc. The level of recovery directly affects the service level provided to the end-user in the event of a network failure. There is a correlation between the level of recovery provided and the cost to the network.

The general description of the functional architecture is applicable for both LSPs and pseudowires (PWs).

This framework applies to general LSP recovery schemes, but also to schemes that are optimized for specific topologies in order to handle protection switching in an ~~cost-~~efficient manner. Recovery schemes for PWs are introduced in Section 6, but the details are for further study and will be addressed in a separate document in the PWE3 working group.

This document takes into account the need for co-ordination of protection switches at multiple layers. This allows an operator to prevent races and allows the protection switching mechanism of one layer to fix a problem before switching at another layer.

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

Generally, network operators aim to provide the fastest, most stable, and the best protection mechanism at a reasonable cost according to the requirements of the customers. The higher the levels of protection, the greater the number of resources consumed and so the higher the likely cost both to the operator and to the customer. It is therefore expected that network operators will offer a wide spectrum of service levels. MPLS-TP-based recovery offers the flexibility to select the recovery mechanism, choose the granularity at which traffic is protected, and also choose the specific types of traffic that are to be protected. With MPLS-TP-based recovery, it is possible to provide different levels of protection for different classes of service, based on their service requirements.

Comment [M29]: The reference model in MPLS-TP OAM only describes the OAM (detection) aspects not the recovery aspects.
GA: Protection requires the relevant information in the OAM to activate protection switching rather than just the availability of OAM

Comment [M30]: Not always true – in a large network 1+1 end to end protection may have lower availability than shared recovery on a set of concatenated domains – shared protection would be lower cost.
GA: This is a subjective statement and it should be deleted

Comment [M31]: PW is end to end only

Comment [M32]: Applies to nested restoration domains as well.

Comment [M33]: Not appropriate to use the term "layer"

Comment [M34]: Support of the tools is mandatory use of the tools is optional

Formatted: Highlight

Formatted: Highlight

Comment [G35]: Not clear. What is meant by higher level of protection. Does this mean higher availability

Formatted: Highlight

Formatted: Highlight

Formatted: Highlight

Comment [M36]: Please define "types of traffic" the draft is about protecting LSPs, links. Use of the term "types of traffic" could imply that recovery is on a per class of service within an LSP.

Comment [G37]: The introduction section is too long and wordy.

2. Terminology and References

The terminology used in this document is consistent with that defined in [RFC4427]. That RFC is, itself, consistent with [G.808.1].

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

Readers should refer to those documents for normative definitions. This document supplies brief summaries of some terms for clarity and to aid the reader, but does not re-define terms.

Comment [M38]: Is RFC4427 or G.808.1 the normative definition for terms that are defined in both?

In particular, note the distinction and definitions made in [RFC4427] for the following three terms.

- Protection: re-establishing end-to-end traffic using pre-allocated resources.
- Restoration: re-establishing end-to-end traffic using resources allocated at the time of need. Sometimes referred to as "repair".
- Recovery: a generic term covering both Protection and Restoration.

Comment [M39]: Is the term "repair" used in this context in this draft or in the more general context of (physical) replacement of a failed component (e.g. circuit pack).

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 refers to the combination of failure detection, localization, and notification mechanisms.
- Failure is used to indicate both signal failure and signal degradation event.
- Trigger indicates any event that may be used to cause an implementation to consider taking protection action.
- The acronym OAM is defined as Operations, Administration and Maintenance consistent with [OAM-SOUP].

Comment [M40]: Defined in MPLS-TP NM framework?

Comment [M41]: In the ITU-T the terms anomaly, defect and failure are used. Recovery is triggered by a defect. Failure is a defect that persists for 2 seconds.

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

3. Requirements for Survivability

MPLS-TP requirements are presented in [RFC5654] and serve as a normative reference for the definition of all MPLS-TP function including survivability. Survivability is presented in [RFC5654] as an important performance parameter ~~critical factor~~ in the delivery of reliable services, and the requirements for survivability are set out using the recovery terminology defined in [RFC4427].

These requirements are summarized below. Reference numbers refer to the requirements as presented in [RFC5654]. Readers should refer to [RFC5654] for the definitive list of requirements which is not replaced or superseded by the list provided here.

3.1. General Requirements

- Protection and restoration mechanisms must be provided (56).
- Recovery techniques should be as similar as possible to those in existing transport networks (56A).
- Point-to-point (P2P) and point-to-multipoint (P2MP) recovery techniques should be the same if possible (56B).
- Recovery must be applicable to links, transport paths, segments, concatenated segments, connections and end-to-end LSPs and PWs (57).
- Recovery objectives must be configurable to meet the SLA objectives of the services offered including rapid (sub-50ms) recovery, protection of all traffic on a path, and protection across multiple domains (58, 59).
- The recovery mechanisms should be applicable to any topology (60).See also Section 3.4.
- Recovery must be coordinated across network layers (61).
- Recovery and reversion must not 'flap' (62).

Note that there is no requirement for support for extra traffic [RFC4427] except in a ring where MPLS-TP must support the sharing of protection bandwidth in a ring by allowing best-effort traffic (108).

3.2. Requirements for Restoration

- The restored and protected paths must be able to share resources (70).

Comment [M42]: How is the term survivability in 5654 related to the term recovery in this draft? Do we need both terms: Either pick one or describe the difference.

Comment [HvH43]: The danger of summarizing is that is allows for editors interpretation, it is better to point only, or quote exactly, it is proposed to mention the requirement numbers only, an dif necessary quote exactly the text.

Comment [G44]: Should the techniques be the same or similar?

Comment [HvH45]: 5456 says identical

Formatted: Highlight

Comment [M46]: Of this draft or 5654?

- Priorities must be available to control the order of restoration and to facilitate preemption during restoration (71, 72).

- Reversion must be supported (73).

3.3. Requirements for Protection

- MPLS-TP data plane protection must operate without regard to payload content (63).

- The following protection schemes must be supported:

  * reversion (64).

  * unidirectional and bidirectional 1+1 protection for P2P (65A, 65B)).

  * unidirectional 1+1 protection for P2MP (65C).

  * bidirectional 1:n protection for P2P (67A).

  * unidirectional 1:n protection for P2MP (67B).

- It must be possible to share protection resources (66). This includes:

  * 1:n mesh recovery should be supported (68).

  * sharing of resources between protection paths ~~that~~ will not be required to protect the same fault (69).

3.4. Requirements for Survivability in Ring Topologies

- MPLS-TP recovery mechanisms may be optimized for specific topologies provided such optimizations interoperate with, and be as similar as possible to, standard techniques to provide~~d~~ end-to-end recovery (91, 100).

- Ring topologies support must include:

  * single ring (92)

  * interconnected rings (93)

  * connection of rings to arbitrary networks (99)

  * logical and physical rings (101)

- Traffic protection in rings must include:
  - unidirectional and bidirectional P2P paths (94)
  - unidirectional P2MP paths (95)
- Ring recovery techniques:
  - must default to bidirectional (102)
  - must support reversion as the default behavior (103)
  - must distinguish (to the operator) trigger mechanisms (104)
  - should protect against multiple failures (106B)
  - must support sharing of protection resources (109)
  - must prevent recovery flapping (107)
- Ring protection mechanism scaling must include:
  - 1+1 and 1:1 protection switching 50 ms from the moment of fault detection in a network with a 16-node ring with less than 1200km of fiber (96)
  - independence from the number of LSPs crossing the ring (97)
  - good performance with increases in the number of transport paths, the number of nodes on the ring, and the number of ring interconnects (98)
- It must be possible to disable protection mechanisms on selected links in a ring (105).
- MPLS-TP recovery mechanisms in a ring must support prioritization of recovery actions arising from different commands or triggers and for different protected entities (106A).

3.5. Triggers for Protection, Restoration, and Reversion

- Triggers must be supported from:
  - lower network layers (74)
  - MPLS-TP OAM (75)

Comment [HvH47]: This is not the same as 5456: 104 The recovery mechanisms in a ring MUST support ways to allow administrative protection switching, to be distinguished from protection switching initiated by other triggers.

Comment [G48]: This does not apply to a ring architecture. This should read that protections switching time shall not exceed 50 ms …

Formatted: Highlight

Formatted: Highlight

Comment [G49]: What does this mean? Switching time?

Formatted: Highlight

Comment [G50]: Text requires clarification. I suggest something that " Events that trigger protection, restoration and reversion" etc.

Formatted: Highlight

* the management plane (76)

* the control plane (if present) (78)

- It must be possible to distinguish trigger sources and to prioritize recovery action requests (77, 79).

3.6. Management Plane Operation

- Support is required for preplanning, pre-calculation, and pre-provisioning of recovery paths and groups of paths (80, 81, 82, 85).

- External commands (controls) must allow the operator to activate~~effect~~, prevent, or test protection switching without triggering~~effecting~~, any recovery operation (83, 84).

- It must be possible to configure all aspects of recovery (86).

- It must be possible to monitor all aspects of recovery (87, 88).

3.7. Control Plane and In-band OAM

- If a control plane is used, it must be possible operate all aspects of recovery (89).

- In-band OAM must support administrative control and protection state coordination (90).

4. Functional Architecture

This section presents an overview of the elements of the functional architecture for survivability within an MPLS-TP network. The intention is to decompose~~break~~ the survivability components into ~~out as~~ separate items so that it can be seen how they may be combined to provide different levels of recovery to meet the requirements set out in the previous section.

4.1. Elements of Control

Survivability is achieved through specific actions taken to repair network resources or to redirect traffic onto paths that avoid failures in the network. Those actions may be triggered automatically by the MPLS-TP network nodes (detecting a network failure), may be enhanced by in-band (i.e. data-plane OAM ~~based~~) OAM fault management or performance monitoring, in-band or out-of-band control plane signaling, or may be under direct the control of an operator.

Comment [YT51]: In 4.1, subsections under it just explain/repeat the backbround of clause 3.5 to 3.7 (or soften requirements in sec 3), therefore they produce nothing valuable. And they are not MPLS-TP specific. I am not sure this subsections are really required.

Comment [M52]: Survivability = recovery?

Comment [G53]: Replace with recovery

Comment [M54]: In which context is the term repair used here? If it is in the 4427 context how is this different from redirecting traffic onto paths that avoid the failure?

Comment [M55]: How? The term “fault management” is normally used in the context of an OSS

Comment [M56]: How would the control plane be aware of the fault if the control plane and forwarding plane are independent (as per the requirements).

Comment [M57]: The operator may prevent recovery or invoke protection restoration for “operational reasons” e.g. to allow routine maintenance. Not clear how an operator would invoke recovery after a failure.

These different options are explored in the next sections.

4.1.1. Manual Control

The survivability behavior of the network as a whole, and the reaction of each LSP when a fault is reported, may be under operator control. That is, the operator may establish own network-wide or local policies that determine what actions will be taken when different failures are reported that affect different LSPs. At the same time, when a service request is made to cause the establishment of one or more LSPs in the network, the operator (or requesting application) may express a required or requested level of service, and this will be mapped to particular survivability actions taken before and during LSP setup, after the 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) onto recovery schemes within the network.

The operator can also be given manual control of survivability actions and events. For example, the operator may perform the following actions:

- inhibit survivability actions
- enable or disable survivability function
- induce the simulation of a network fault or force a switchover from a working path to a recovery path (for network optimization purposes with minimal disturbance of services, 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.

4.1.2. Failure-Triggered Actions

Survivability actions may be directly triggered by network failures. That is, the device that detects the failure (for example, detection of Loss of Light on an optical interface, a failure to receive an OAM Continuity message, or a reception of OAM Alarm Report) may immediately perform a survivability action. Recall that the term "failure" is used to represent both signal failure and signal degradation.

Comment [M58]: Should reference G.808.1 for the description of these actions.

Comment [M59]: This should be related to the recovery domain concept - policies apply only within the context of a domain. A policy on an end to end path cannot impact the policy applied to a segment (that may be in a different administrative domain.

Comment [G60]: The term failures should be replaced by defects

Formatted: Highlight

Comment [G61]: I assume actions are different from functions below. The use of the term survivability is used loosely to mean protection/restoration/recovery

Formatted: Highlight

Comment [M62]: How is this different from inhibiting or removing the inhibition?

Comment [M63]: Exerciser in G.808.1?

Comment [M64]: Manual/forced switch in G.808.1?

Formatted: Highlight

Formatted: Highlight

Comment [G65]: This should be covered by external commands.

Formatted: Highlight

Comment [M66]: Presumably this would only occur if (automatic) recovery was not provided for the failed service. In which case the action would be to reroute and reprovision the path (from the OSS). Is this recovery within the context of this draft?

Comment [M67]: Should reference G.808.1 for the definition of these conditions.

Comment [M68]: How would a termination in MPLS-TP or PW layer be aware of a "loss of light" – it may get a server fail defect indication.

Comment [M69]: What is an OAM alarm report?

This behavior can be subject to management plane or control plane control, but does not require any messages exchanges in any of the management plane, control plane, or data plane to trigger the recovery action - it is directly triggered by defect events in the data plane ~~stimuli~~.

Note, however, that coordination of recovery actions between the edges of the recovery domain may require message exchanges for some qualitative levels of recovery or when performing a bidirectional recovery action.

4.1.3. OAM Signaling

OAM signaling refers to message exchanges that are in-band or closely coupled to the data channel. Such messages may be used to detect and isolate faults or indicate a degradation in the operation of the network, but in this context we are concerned with the use of these messages to control or trigger survivability actions.

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 are not under management plane control. The control plane may also be used to detect, isolate, and communicate network failures 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 resource-availability information. In this way, "graceful shutdown" of resources may be effected by withdrawing them, and this can be used as a stimulus to survivability action in a similar way to the reporting or discovery of a fault as described in the previous sections.

4.2. Elements of Recovery

This section describes the elements of recovery. These are the quantitative aspects of recovery; that is the pieces 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] a mapping is provided.

[Comment: the following Sections:
4.2.1 Span Recovery
4.2.2 Segment Recovery
4.4.1 Link Level protection
4.4.3 Protection Tunnels
All appear to be attempting to describe the same construct.

Comment [M70]: The reaction of the protection (or recovery) state machine to the inputs it receives is independent of the management plane and control plane. The management plane may provide some input (e.g. lockout; forced switch). The control plane cannot influence the protection state machine.

Comment [M71]: Please explain this.

Comment [M72]: In G.808.1 this is refered to as APS signalling. OAM and APS may use an Ach but they will be different instances.

Comment [M73]: Isolation is not required to trigger recovery

Comment [M74]: Degradation is normally determined by the processing of AOM messages (e.g. to determine loss or delay defects)

Comment [M75]: OAM can only trigger recovery, APS signalling is used for control/coordination.

Comment [M76]: If we have control plane/forwading plane independance (as required) the control plane cannot detect forwarding plane faults.

Comment [M77]: Does this support “make before break i.e. a new path is established before the old path is withdraw. Where is this described – please provide a normative reference.

Comment [M78]: Insert 4.5 and 4.6 here

We should use the term "tunnel" to refer to a server aggregation construct that carries one or more client LSPs. The tunnel may be protected i.e. we have a working tunnel and standby tunnel. If the working tunnel fails then the client LSP's are carried by the standby tunnel.]

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 although 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, however, a TE link may be provided by:

- 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 a collection of data links that, together, provide data link layer protection scheme.

Thus, span recovery may be provided by:

- 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
- re-routing 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 is described in Section 4.2.2.

[RFC5654] refers to a span as a "link".

4.2.2. Segment Recovery

An LSP segment is one or more continuous hops on the path of the LSP. [RFC5654] defines two terms. A "segment" is a single hop on the path of an LSP, and a "concatenated segment" is more than one hop on 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 out of multi-segment PW (MS-PW) that is set up between two PE devices (i.e. T-PE and S-PE, S-PE and S-PE, or S-PE and T-PE). As indicated in Section 1, the recovery of PWs and PW segments is for further study and will be described in a separate document in the PWE3 working group. See also Section 6 of this document.

Comment [M79]: How is this different from link recovery

Comment [YT80]: In 4.2.1 and 4.2.2 there seems to be the overlapping item between single (one hop) segment and {span or link.
For those clause, it could be better restructure
- link
- single segment
- multiple segments (i.e. TCM)

Comment [HvH81]: So, what is a span? A TE link or a data link.

Comment [HvH82]: Why not use link in this draft as well to be consistent?

Comment [M83]: Segments must be independent each “domain” has a recovery segment:

How about dual node interconnect between domains?

LSP segment recovery involves redirecting of traffic at one end of a segment of an LSP onto an alternate path to the other end of the segment. According to the required level of recovery (described in Section 4.3), this redirection may be onto a pre-established LSP segment, through re-routing of the protected segment, or by tunneling the protected LSP through a "bypass" LSP. For details on recovery mechanisms, see Sections 4.4 and 4.5 below.

Note that protecting an LSP against the failure of a node requires the use of segment recovery, while a link could be protected using span or segment recovery.

4.2.3. End-to-end Recovery

End-to-end recovery is a special case of segment recovery where the protected LSP segment is the whole of the LSP. End-to-end recovery may be provided as link-diverse and/or node-diverse recovery where the recovery path shares no links and/or no nodes with the recovery path.

OLD:

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

NEW:

Note that in the absence of Shared Risk Link Group (SRLG) restrictions, node-diverse paths are link-diverse, and that full, end-to-end node-diversity is required to guarantee recovery.

4.3. Levels of Recovery

This section describes the qualitative levels of survivability function that can be provided. The level of recovery offered has a direct effect on the service level provided to the end-user in the event of a network fault. This will be observed as the amount of data lost when a network fault occurs, and the length of time to recover connectivity.

In general there is a correlation between the service level (i.e. the rapidity of recovery and reduction of data loss) and the cost to the network; better service levels require pre-allocation of resources to the recovery paths, and those resources cannot be used for other purposes if high quality recovery is required. Thus, 'cost' in this case may be measured as the financial cost of providing resources for the recovery scheme, or the financial loss from dedicating resources to the recovery scheme such that they cannot be used to draw new revenue.

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

4.3.1. Dedicated Protection

In dedicated protection, the resources for the recovery LSP are pre-assigned for use only by the protected service. This will clearly be

Comment [HvH84]: In 1+1 it is copying

Comment [HvH85]: The source side

Formatted: Highlight

Comment [M86]: Not always. For example the server layer paths may be multiplexed onto the same cable/fiber or wavelength at some remote site.

Formatted: Highlight

Comment [M87]: Should use the term allocated since in 1+1 traffic is always sent on both paths thus the protection resources are always in use.

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 in the bypass tunnel recovery mechanism (see Section 4.4.3) resources may also be dedicated to protecting a specific service. In some cases (one-for-one protection) the whole of the bypass tunnel may be dedicated to provide recovery for a specific LSP, but in other cases (such as facility backup) a subset of the resources of the bypass tunnel may be pre-assigned for use to recover a specific service. However, as described in Section 4.4.3, the bypass tunnel approach can also be used for shared protection (Section 4.3.2), to carry extra traffic (Section 4.3.3), or without reserving resources to achieve best-effort recovery.

4.3.2. Shared Protection

In shared protection, the resources for the recovery LSPs of several services are shared. These may be shared as 1:n or m:n, and may be shared on individual links, on LSP segments, on PW segments, or on end-to-end transport path (LSP or PW). Note that as indicated in Section 3 and [RFC5654], m:n recovery is not required in MPLS-TP.

Where a bypass tunnel is used (Section 4.4.3), the tunnel might not have sufficient resources to simultaneously protect all of the paths to which it offers protection, so that if they were all affected by network failures at the same time, they would not all be recovered.

Shared protection is a trade-off between ~~expensive~~ network resources being dedicated to protection that is not required most of the time, and the risk of unrecoverable services in the event of multiple network failures. There is also a trade-off between rapid recovery (that can be achieved with dedicated protection, but which is delayed by message exchanges in the management, control, or data planes for shared protection) and the reduction of network cost by sharing protection resources. These trade-offs may be somewhat mitigated by using m:n for some value of m != 1, and by establishing new protection paths as each available protection path is put into use.

4.3.3. Extra Traffic

A way to utilize network resources that would otherwise be idle awaiting use to protect services, is to use them to carry other traffic. Obviously, this is not practical in dedicated protection (Section 4.3.1), but is practical in shared protection (Section 4.3.2) and bypass tunnel protection (Section 4.4.3).

When a network resource that is carrying extra traffic is required for protection, the extra traffic is disrupted - essentially it is

Comment [M88]: In 1:1 the resources are assigned but not allocated i.e. the LSP is known but traffic is not placed on the protection path until a bridge is invoked (e.g. as the result of a fault). The resources are only allocated when the traffic is actually being carried by the protection chanel.

Comment [M89]: How is this relevant since extra traffic is not supported in MPLS-TP

Comment [M90]: What is shared? The label on the path or the resources (bandwidth) to support the path. Assign vs. allocate.

Comment [M91]: Sharing of resources is always per link, sharing may occur in multiple concatenated links on a LSP segment.

Comment [M92]: In this case would the assignment be active but the resource allocation policy (based on QoS) may result in a high rate of discard. Or, is the recovery action not taken i.e. the bypass tunnel label would not be activated.

Comment [M93]: How does this relate to the earlier statements that protection is 10's of ms restoration is 100's ms to seconds?

Comment [M94]: How is this relavent if m:n is not supported?

Comment [M95]: Why do we need this section if Extra Traffic is not supported?

pre-empted by the recovery LSP. This may require some additional messages exchanges in the management, control, or data planes, with the consequence that recovery may be delayed somewhat. This provide an obvious trade-off against the cost reduction (or rather, revenue increase) achieved by carrying extra traffic.

Note that in MPLS-TP support for extra traffic is not required except in ring topologies (Section 3 and [RFC5654].

4.3.4. Restoration and Repair

This section refers to LSP restoration and repair. Restoration for PWs is for further study and will be described in a separate document in the PWE3 working group (see also Section 6). If resources are not pre-assigned for use by the recovery LSP, the recovery LSP must be established "on demand" when the network failure is detected and reported, or upon instruction from the management plane.

Restoration represents the most cost-effective use of network resources as no resources are tied up for specific protection usage. However, restoration requires computation of a new path and activation of a new LSP (through the management or control plane). These steps can take much more time than is required for 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 the use of LSP setup priorities so that recovery LSPs can pre-empt other low priority LSPs.

Additionally, when a network failure occurs, multiple LSPs may be disrupted by the same event. These LSPs may have been established by different Network Management Stations (NMSs) or signaled by different head-end MPLS-TP nodes, and this means that multiple points in the network will be trying to compute and establish recovery LSPs at the same time. This can lead to contention within the network meaning that some recovery LSPs must be retried resulting in even slower recovery times for some services.

Both hard and soft LSP restoration may be supported. In hard LSP restoration, the resources of the LSP are released before the full establishment of the recovery LSP (i.e. break-before-make). In soft LSP restoration, the resources of the LSP are released after the full establishment of an alternate LSP (i.e. make-before-break).

Note that the restoration resources may be pre-calculated and even

Comment [M96]: Please explain the difference between restoration and repair

Comment [YT97]: the clarification for adding "repair" is required since not in RFC4427.

Comment [M98]: Recovery?

Comment [M99]: Even if the resources are pre-assigned as described in the paragraph above?

Comment [M100]: Do we have the case where resources are not release even after a recovery path is established.
How is the requirement for revertive restoration met if hard or soft restoration is used as defined in this paragraph?

Comment [M101]: Is this different from the pre assignment described above?

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. The pre-planning may happen periodically to have the most accurate information about the available resources in the network.

4.3.5. Reversion

When a service has been recovered so that traffic is flowing on the recovery LSP, the faulted network resource may be repaired. The choice must be made about whether to redirect the traffic back on to the original working LSP, or to leave it where it is on the recovery LSP. These behaviors are known as "revertive" and "non-revertive", respectively.

In "revertive" mode, care should be taken to prevent frequent operation of the recovery operation due to an intermittent defect. Therefore, when the failure condition of a recovery element has been handled, a predetermined period of time should elapse before normal data traffic is redirected back onto the original working entity. It should be possible for an operator to configure this period of time per LSP. A default value should be defined.

4.4. Mechanisms for Recovery

The purpose of this section is to describe in general (MPLS-TP non-specific) terms the mechanisms that can be used to provide protection. 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 segment are for further study and will be described in a separate document in the PWE3 working group (see also Section 6).

4.4.1. Link-Level Protection

Link-level protection refers to two paradigms: (1) where the protection is provided in a lower network layer, and (2) the 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 to link-level protection to protect against a failure of the edge nodes.

Link-level protection offers the following levels of protections:

Comment [HvH102]: It would be good to describe why revertive operation is required

Comment [HvH103]: defective

Comment [M104]: i.e. defective hardware is replaced

Comment [G105]: The terminology used is not consistent through the document

Comment [M106]: Reference wait to restore in G.808.1

Comment [YT107]: the title should be Mechanism for "Protection" since no refering to restoration in it.

Comment [M108]: Does this description also apply to restoration?

Comment [M109]: How are these different for the perspective of the client?

- 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 onto the protecting entity. In this scenario, all LSPs carried over the 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 carried over a given 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 which may be preempted when protection switching occurs.

As with recovery in layered networks, a protection mechanism at the lower layer needs to be coordinated with protection actions at the upper layer in order to avoid race conditions. In general, this is arranged to allow protection actions to be performed in the lower layer before any attempt is made to perform protection actions in the upper layer.

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 link level and use these channels to monitor the health of the MPLS-TP link and 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 same network layer as the protected LSP either for the entire end-to-end LSP or for a segment of the LSP. In this case, hierarchical LSPs are not used - compare with Section 4.4.3.

Different levels 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 there is a failure condition on the working entity, the traffic is switched onto the protection entity. Dedicated protection may be performed using 1:1 or 1+1 protection schemes. When the failure condition is eliminated, the traffic may revert to the working entity. This is subject to

Comment [YT110]: the protecion diagrams of (1) and (2) should be clarified in terms of scheme as described such in 4.5.1.1.1.

Comment [M111]: Not true if its provided by a server layer

Formatted: Highlight

Comment [M112]: Applies to both layer and nested restoration domains.

Comment [G113]: Terminology used is different from other sections. Recovery entity is used rather protection entity.

local configuration.

- Shared protection, where one or more protection entity 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 a hierarchical LSP that is pre-provisioned in order to protect against a failure condition along a network segment, which may affect one or more LSPs that transmit over the network segment.

When there is a failure condition in the network segment, one or more of the protected LSPs are switched over at the ingress point of the network segment and transmitted over the protection tunnel. The way to realize this is using label stacking. Label mapping may be an option as well.

Different levels of protection may be provided:

- Dedicated protection, where the protection tunnel has resource reservations sufficient to provide protection for all protected LSPs without service degradation.

- Shared protection, where the protection tunnel has resources to protect some of the protected LSPs, but not all of them simultaneously.

4.5. Protection in Different Topologies

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

It is required that such mechanisms interoperate with the mechanisms defined for arbitrary topologies to allow end-to-end protection and to allow consistent protection techniques to be used across the whole network.

This section describes two different topologies and explains how

Comment [M114]: How is this different from link protection

Comment [G115]: This is a new term. Is this a sub network?

Comment [M116]: Is the failure detected by the network segement or the LSP on the segment. Is the protection invoked for the segment or for each LSP using the segnement?

Comment [M117]: How does the protection tunnel decide which LSPs to protect or does it attempt to protect all and use QoS to give preference to the traffic from the priority LSPs.

Comment [M118]: If they are in independent recovery domains what aspects need to interoperate?

recovery may be markedly different in those different scenarios. It also introduces the concept of a recovery domain and shows how end-to-end survivability may be achieved through a concatenation of recovery domains each providing some level of recovery in part of the network.

4.5.1. Mesh Networks

Linear protection provides a fast and simple protection switching mechanism and fits best in mesh networks. It can protect against a failure that may happen on a~~n~~ node, a span, an LSP segment, or an end-to-end LSP. Linear protection provides a clear indication of the protection status.

Linear protection operates in the context of a Protection Domain. A Protection Domain is a special case of a Recovery Domain [RFC4427] that applies to the protection function. A Protection Domain is composed of the following architectural elements:

- A set of end points which reside at the boundary of the Protection Domain. In this simple case of 1:n or 1+1 P2P protection, exactly two endpoints reside at the boundary of the Protection Domain. In each transmission direction one of the end points is referred to as a source and the other one is referred to as a sink. In the case of unidirectional P2MP protection, three or more endpoints reside at the boundary of the Protection Domain. One of the endpoints is referred to as source/root and the other ones are referred to as sinks/leaves.

- A Protection Group which consists of a working (primary) path and one or more recovery (backup) paths which run between the endpoints of the Protection Domain. In order to guarantee protection in all situations, a dedicated recovery path should be pre-provisioned to protect against a failure of a working path (i.e. 1:1 or 1+1 protection schemes). Also the working and the recovery paths should be disjoint, i.e., the physical routes of the working and the recovery paths should have complete physical diversity.

Note that if the resources of the protection path are less than~~t~~ 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 such a case, the protection path might not have sufficient resources to simultaneously protect all of the working paths that may be affected by fault conditions at the same time.

Comment [M119]: This description should be moved to the beginning of the draft.

Comment [YT120]: it seems that only linear protection schemes are applicable and that other mechanisms for protection such as node, link (Span or single segment), and (multi hop) segments are out of scope before description. These others mechanism should be considered even if they might be FFS.

Comment [G121]: Why linear protection is so special to mesh networks?

Comment [M122]: Please explain this – linear protect is between a pair of points so how is a mesh relavent?

Comment [G123]: I do not see the difference between the two

Comment [M124]: What is special? The domain construct also applies to restoration.

Comment [G125]: Terminoloy change again. Recovery vs Protection.

Comment [M126]: What is being shared – the resources of the links or the label for protection.

Comment [M127]: In 1:n the assignment of protection is shared by multiple working paths and hence by definition it can only protect a single working path.

For P2P paths, both unidirectional and bidirectional protection switching is supported. In bidirectional protection switching, in the event of failure, the recovery actions are taken in both directions (even when the fault is unidirectional). This requires some level of ~~synchronization~~ coordination of the recovery state between the endpoints of the protection domain.

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

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

4.5.1.1. Protection Schemes in Mesh Topologies

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

4.5.1.1.1. 1:n Linear Protection

In the 1:1 scheme, a recovery path is allocated to protect against a failure ~~of~~ or degradation in a working path. As described above, in order to guarantee protection, the recovery entity should support the full capacity and bandwidth, but it may be degraded from the normal working entity.

Figure 1 presents 1:1 protection architecture. In normal conditions the data traffic is transmitted over the working entity and the recovery entity is in an idle state. Normal conditions are defined when there is no failure or degradation on the working entity and there is no administrative configuration or requests that cause traffic to transmit over the recovery entity.

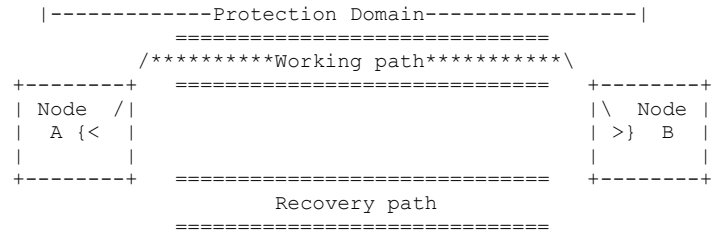

Figure 1: 1:1 protection architecture

Upon a fault condition (failure or degradation) along the working

Comment [M128]: The mesh is reduced to a point to point topology...

Comment [YT129]: section following from 4.5.1.1 should be made independent from 4.5 since this is not mesh specific

Comment [G130]: This is not clear. Does it mean that if the protection channel does not have the full capacity, the signal will be degraded.

Comment [M131]: How would it be degraded?

Comment [M132]: Data plane OAM should be active on the recovery entity to ensure that it is operable.

entity or a specific administrative request, the traffic is switched over to the recovery entity.

Note that in the non-revertive behavior (see section 4.3.5), data traffic can be transmitted over the recovery entity also in normal conditions. This can happen after the condition(s) causing the switchover has/have been cleared.

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

In bidirectional protection switching, both ends of the protection domain switch to the recovery entity (even when the fault is unidirectional). This requires a protocol to try and synchronize the protection state between the two end points of the Protection Domain.

When there is no failure, the resources of the idle entity may be used for less priority traffic. When protection switching is performed, the less priority traffic may be pre-empted by the protected traffic.

In the general case of 1:n linear protection, one recovery entity is allocated to protect n working entities. The Protection entity might not have sufficient resources to simultaneously protect all of the Working entities that may be affected by fault conditions at the same time.

In case of failures along multiple working entities, priority should be set as to which entity is protected. The protection states between the edges of the Protection Domain should be fully synchronized to ensure consistent behavior. As explained above in section Revertive behavior is recommended when 1:n is supported.

4.5.1.1.2. 1+1 Linear Protection

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

As depicted in figure 2, data traffic is copied ~~at~~ and fed at the source to both the working and the recovery entities. The traffic on the working and the recovery entities is transmitted simultaneously to the sink of the Protection Domain, where the selection between the

Comment [M133]: IN this case is traffic left on the "old" working path. If so this has become 1+1 following the first protection event.

Comment [M134]: Use the description in G.808.1

Comment [M135]: What is shared the label of the path or just the resources supporting the path.

Comment [M136]: How is pre-emption achieved. The label for the lower priority traffic is withdrawn or QoS causes discard?

Comment [M137]: See comment about this above

working and recovery entities is made (based on some predetermined criteria).

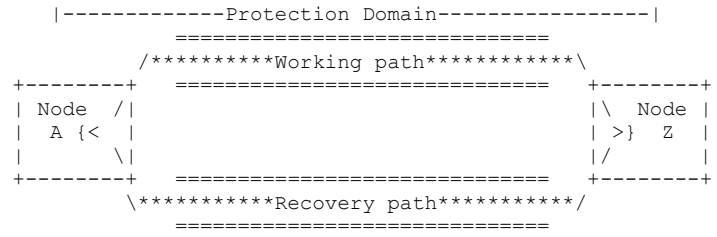

Figure 2: 1+1 protection architecture

Note that control traffic between the edges of the Protection Domain (such as OAM or control protocol to synchronies the protection state, etc.) may be transmitted on a different entity than 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 recovery state between the protection controllers at both ends of the protection domain. In 1+1 bidirectional protection switching, there is a need for a protocol to coordinate the protection state between the edges of the Protection Domain.

In both protection schemes traffic is restored to the working entity after the condition(s) causing the switchover has/have been cleared. To avoid frequent switching in case of intermittent failures when the network is not stabilized, traffic is not switched back to the working entity before the Wait-to-Restore (WTR) timer has expired.

Comment [M138]: Non-revertive operation?

4.5.1.1.3. P2MP linear protection

Linear protection may apply to protect unidirectional P2MP entity using 1+1 protection architecture. The source/root MPLS-TP node bridges the user traffic to both the Working and Protected entities. Each sink/leaf MPLS-TP node selects the traffic from one entity based on 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 requires the sink/leaf MPLS-TP nodes to notify the source/root MPLS-TP node of any fault condition. This required also 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 recovery transport path to transfer the traffic.

Note that such a mechanism does not exist and its exact behavior ~~if~~is for further study.

4.5.1.2. Triggers for the Linear Protection Switching Action

The protection switching may be performed when:

- A fault condition ('failed' or 'degraded') is declared on the working entity and is not declared on the recovery entity. Proactive in-band OAM CC$amp;V (Continuity and Connectivity Verification) monitoring of both the working and the recovery entities may be used to enable the fast detection of a fault condition. For protection switching, it is common to run a CC$amp;V every 3.33ms. In the absence of three consecutive CC$amp;V messages, a fault condition is declared. In order to monitor the working and the recovery entities, an OAM Maintenance Entity should be defined for each of the entities. OAM indications of fault conditions should be provided to the edges of the Protection Domain which are responsible for the protection switching operation. Input from OAM performance monitoring indicating degradation in the working entity may also be used as a trigger for protection switching. In the case of degradation, switching to the recovery entity is needed only if the recovery entity can guarantee better conditions.

- An indication is received from a lower layer server that there is a network failure.

- An external operator command is received (e.g. 'Forced Switch', 'Manual Switch'). For details see Section 5.1.2.

- A request to switch over is received from the far end. The far end may initiate this request for example when it gets an administrative request o switch over, or when bidirectional 1:1 protection switching is supported and there was a fault that could be detected only by the far end, etc.

As described above, in some cases an attempt should be done to coordinate the protection states between the end points of the

Comment [G139]: Perhaps it should say Events that triggers protection switching

Comment [M140]: The triggers should be independent of the type of recovery being used – linear protection, ring protection, restoration….

Comment [HvH141]: Working can have SF and protection SD

Comment [M142]: Should the defect detection process be described in the OAM framework instead of this document.

Comment [M143]: The framework should describe exactly what coordination MUST be performed and what action is taken if the required coordination fails.

Protection Domain. Control message should be exchanged between the edges of the Protection Domain to synchronize the protection state of the edge nodes. The control messages can be delivered using in-band data-plane driven control protocol or a control plane based protocol.

In order to achieve 50ms protection switching it is recommended to use in-band data-plane driven signaling protocol to coordinate the protection states. An in-band data-plane PSC (Protection State Coordination) protocol is defined in [MPLS-TP-Linear-Protection] for this purpose. This protocol is also used to detect mismatches between the configuration provisioned at the ends of the Protection Domain.

As described below in section 5.5, GMPLS already defines procedures and messages' elements to synchronize the protection states between the edges of the protection domain. These procedures and protocols messages are specifies in [RFC4426], [RFC4872] and [RFC4873]. However, these messages lack the capability to synchronize the revertive/non-revertive behavior and the consistency of configured timers at the edges of the Protection Domain (timers such as Wait to Restore (WTR), Hold-off timer, etc.).

4.5.1.3. Applicability of linear protection for LSP segments

In order to implement data-plane based linear protection on LSP segments, there is a need to support the MPLS-TP architectural element PST (Path Segment Tunnel). Maintenance operations (e.g. monitoring, protection or management) engage with a transmission of messages (e.g. OAM, Protection Path Synchronization, etc.) in the maintained domain. According to the MPLS architecture which is defined in [RFC3031], such messages can be initiated and terminated at the edges of a path where push and pop operations are enabled . ~~As an exception, these messages may be terminated at an intermediate node when the TTL value is expired~~. In order to support the option to monitor, protect and manage a portion of an LSP, a new architectural element is defined, Path Segment Tunnel (PST). A Path Segment Tunnel is an LSP which is basically defined and used for the purposes of OAM monitoring, protection or management of LSP segments. PST makes use of the MPLS construct of hierarchical nested LSP which is defined in [RFC3031].

For linear protection operation, PSTs should be defined over the working and recovery entities between the edges of a Protection Domain. OAM and PSC messages can be initiated at the edge of the PST and sent to the peer edge of the PST. Note that these messages are sent over G-ACH channels, within the PST and use two labels stack, the PST label at the bottom of stack and the G-ACH label.

Comment [M144]: APS signalling should be carried by the Ach and MUST be independent of the control plane.

Comment [M145]: Is the Ach inband in this context?

Comment [G146]: Protection switching should operate independently from control or management planes. APS should be autonomous.

Comment [M147]: MPLS-TP must operate in the absence of a control plane. Also note that a control plane may be used for configuration but the availability may not be adequate to support recovery. Also if we need extensions why should we consider using them in MPLS-TP

Comment [M148]: True, but is this relevant for a protection domain?

Comment [M149]: GAL?

The end-to-end traffic of the LSP, including data-traffic and control traffic (OAM, PSC, management and signaling messages) is tunneled within the PSTs by means of label stacking as defined in [RFC3031].

The mapping between an LSP and a PST can be 1:1 which is similar to the ITU-T Tandem Connection element which defines a sub layer corresponding to a segment of a path. The mapping can also be 1:n to allow scalable protection of a set of LSPs' segments traversing the portion of the network in which a Protection Domain is defined. Note that each of these LSPs can be initiated or terminated at different endpoints in the network, but they all traverse the Protection Domain and share similar constraints (such as requiremtns for QoS, terms of protection ,etc.). In case of 1:n mapping PSTs can be referred to also as TE links.

Comment [M150]: Why must they have the same QoS?

Comment [M151]: Where is this term used within MPLs-TP?

Note that in the context of segment recovery, the PSTs serve as the working and protection entities.

4.5.2. Ring Networks

Several Service Providers have expresses a high level of interest in operating MPLS-TP in ring topologies and require a high level of survivability function in these topologies.

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

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

2. Optimization of resource consumption around the ring, like the number of labels needed for the protection paths that cross the network, the total bandwidth needed in the ring to ensure the protection of the paths, etc.

[RFC5654] introduces a list of requirements on ring protection that cover the recovery mechanisms need to protect traffic in a single ring and 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]. The principles for the development of the mechanisms should be:

1. Reuse existing procedures and mechanisms for recovery in ring topologies as along as their performance is as good as new potential mechanisms.

2. Ensure complete interoperability with the mechanisms defined for arbitrary topologies to allow end-to-end protection.

4.5.3. Recovery ~~Protection and Restoration~~ Domains

Protection and restoration are performed in the context of a recovery domain. A recovery domain is defined between two or more recovery reference endpoints which are located at the edges of the recovery domain and bounds the element on which recovery can be provided (as described in section 4.2 above). This element can be end-to-end path, a portion of a path or a span.

The case of an end-to-end path can be observed as a special case of a portion of a path, and the ingress and the egress LERs serve as the recover reference end-points.

In this simple case of a P2P protected entity, exactly two endpoints reside at the boundary of the Protection Domain. An LSP can enter through exactly one reference endpoint and exit the recovery domain through another reference endpoint.

In the case of unidirectional P2MP, three or more endpoints reside at the boundary of the Protection Domain. One of the endpoints is referred to as source/root and the other ones are referred to as sinks/leaves. An LSP can enter the recovery domain through the root point and exit the recovery domain through the leaves points.

The recovery mechanism should restore interrupted traffic due to a facility (link or node) fault within the recovery domain. Note that a single link may part of several recovery domains. If two recovery domains have any links in common, then one recovery domain must be contained with the other. This can be referred to as nested recovery domains. ~~However~~ The boundaries of recovery domains may be coincident, however they must not ~~overlap~~intersect.

Note that the edges of a recovery domain are ~~not protected and unless contained in another recovery domain, they form~~ a single point of failure.

A recovery group is defined within a recovery domain and it consists

Comment [M152]: Move up to the start of 4.2

Comment [M153]: Segment?

Comment [M154]: Segment?

Comment [M155]: Could be contained with a coincident boundary which would offer no protection.

Comment [M156]: Does this only apply to protection? Is so make an explicit scoping statement.

of a working (primary) entity and one or more recovery (backup) entities which reside between the endpoints of the recovery Domain. In order 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.

The method used to monitor the health of the recovery element is outside the scope of this draft. ~~unimportant, provided that t~~ The endpoints which are responsible for the recovery action ~~receive the information on~~MUST be aware of its condition. The condition of the recovery element may be 'OK', 'failed', or 'degraded'.

When the recovery operation is triggered by an OAM FM or PM indication, an OAM Maintenance Entity Group is defined for each of the working and protection entities.

Comment [M157]: Its not defined before the trigger?

The recovery entities and functions in a recovery domain can be provisioned 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 5.1.

Comment [M158]: The control plane may configure protected connections or perform restoration. The control plane cannot configure the boundaries or policies of a recovery domain.

When a control plane is used, the ingress LERs may communicate with the recovery reference points requesting protection or restoration across a recovery domain. For details, see section 5.5.

Comment [M159]: The ingress LER is by definition at a recovery reference point. Faults are normally detected by los of CC/CV at the egress of the domain. Thus the egress LER requests the ingress LER to invoke restoration (source rerouting).

4.6. Recovery in Layered Networks

Comment [M160]: Move with 4.5

In multi-layer or multi-region networking, recovery may be performed at multiple layers or across cascaded recovery domains.

Comment [M161]: Layer and regions are independent concepts. It would be more appropriate to discuss nested/concatenated regions (domains) in the previous section.

The MPLS-TP recovery mechanism must ensure that the timing of recovery is coordinated in order to avoid races, and to allow either the recovery mechanism of the server layer to fix the problem before recovery takes place at the MPLS-TP layer, or to allow an upstream recovery domain to perform recovery before a downstream domain. In inter-connected 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 failure occurred.

A hold-off timer is required to coordinate the timing of recovery at multiple layers or across cascaded 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 failure condition in the MPLS-TP layer should not immediately trigger the recovery process if the hold-off timer is set to a value other than zero. The hold-off timer should be started and, on expiry, the recovery element should be checked to determine whether the failure 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.

Reference should be made to [RFC3386] that presents the near-term and practical requirements for network survivability and hierarchy in current service provider environments.

Comment [M162]: When is the timer started – when a detect is reported?

Comment [G163]: I thought setting the hold-off timer to zero will trigger protection in higher layer rather than the lower layers.

Comment [M164]: Is the hold off implemented by each (layer) recovery mechanism (preferred) or by the deftect processing in the server layer?

Comment [M165]: Please explain how 3386 is this relevant in the context of a transport network?

4.6.1. Inherited Link-Level Protection

Where a link in the MPLS-TP network is formed from 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 level 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 links that themselves have inherited protection.

By classifying the server layer links in the MPLS-TP network according to the level of ~~underlying~~ protection that they have, it is possible to compute an end-to-end path in the MPLS-TP network that uses only links with a specific or better level of inherited protection. This means that the end-to-end MPLS-TP LSP can be protected at the level necessary to conform with the SLA without the need to provide any additional protection in the MPLS-TP layer. This saves complexity and network resources, and reduces issues of protection switching coordination.

Where the requisite level of inherited protection is not available on all segments along the whole path in the MPLS-TP network, ~~it can be "topped up" using protection in the~~ MPLS-TP layer~~. S~~ segment protection ~~would~~ may be used for these segments. ~~particularly suitable.~~

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 any protection in the MPLS-TP or PW layer to handle node failure.

4.6.2. Shared Risk Groups

When an MPLS-TP protection scheme is established, it is ~~essential~~desirable that the working and protection paths do not share resources in the network. If this is not achieved, a single failure may affect both the working and the protection path with the result that the traffic cannot be delivered - it was, in fact, not protected.

Note that this restriction does not apply for restoration as this takes place after the fault has arisen meaning that the point of failure can be avoided.

When planning a recovery scheme it is possible to select paths that use diverse links and nodes within the MPLS-TP network using a topology map of the ~~network~~MPLS-TP layer. 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. And 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. The working and protection paths can be planned to be not only node and link diverse, but to not use any resources from the same shared risk groups.

4.6.3. Fault Correlation

TBD. This is about correlating multiple faults from the lower layers to observe that they all represent the same fault in the MPLS-TP layer.

5. Mechanisms for Providing Protection of MPLS-TP LSPs

This section describes the existing mechanisms available to provide protection of LSPs within MPLS-TP networks, and highlights areas where new work is required. It is expected that, as new protocol extensions and techniques are developed, this section will be updated

Comment [M166]: It may not always be possible to restore since some (major) failures can isolate segments of a network.

Comment [M167]: If the topology is of the network the diversity of the resources should be evident. If the topology is of the network the diversity of the resources should be evident.

Comment [M168]: Important but how does this impact recovery

Comment [YT169]: I am not sure why we need the section 5 as described in details. Especially, section 5.2. to 5.4 are rather related to OAM fwk regarding to protection though the title of head of section 5 is "mechanism". And 5.5 is only saying c-plane aspects and it does not touch MPLS-TP LSPs as titled the head of section 5.

to convert the statements of required work into references to those protocol extensions and techniques.

5.1. 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 request for manual control of protection switching).

The management plane may be used to configure the recovery domain by setting the reference endpoints points (which controls 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 a 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 between fault conditions and 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 OAM PSC (protection State Coordination) which is specified in [MPLS-TP-Linear-Protection] may be used as an in-band (i.e. data plane-based) control protocol to coordinate the protection states between the endpoints of the recovery domain and to check consistency of configured parameters (such as timers, revertive/ non-revertive behavior, etc.)

It should also be possible for the management plane to monitor the recovery status.

Comment [M170]: Also used to configure restoration.

Comment [G171]: If there is a mismatch in the end points configuration then this defect should be flagged

Comment [G172]: The MP receivers reports rather than functionally monitoring the recovery status.

5.1.1. Configuration of Protection Operation

In order to implement the protection switching mechanisms, the following entities and information should be configured and provisioned:

- The endpoints of a recovery domain. As described above, these endpoints bound the element of recovery for 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, in order to guarantee protection, the paths of the working entity and the recovery entities should have complete physical diversity.

- As defined in section 4.5.2, in order to implement data-plane based LSP segment recovery, there is a need to support the MPLS-TP architectural element PST (Path Segment Tunnel), since related control messages (e.g. for OAM, Protection Path Synchronization, etc.) can be initiated and terminated at the edges of a path where push and pop operations are enabled. PST is an end-to-end LSP which corresponds in this context to the recovery entities (working and protection) and makes use of the MPLS construct of hierarchical nested LSP which is defined in [RFC3031]. OAM and PSC messages can be initiated at the edge of the PST and sent to the peer edge of the PST, over G-ACH. There is a need to configure the related PSTs and map between the LSP segment(s) being protected and the PST. The mapping can be 1:1 or 1:N to allow scalable protection of a set of LSPs' segments traversing the portion of the network in which a Protection Domain is defined. Note that each of these LSPs can be initiated or terminated at different endpoints in the network, but 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.

5.1.2. External Manual Commands

The following external, manual commands may be provide for manual control of the protection switching operation. These commands apply to a protection group and 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.

Comment [G173]: To have independent physical paths rather than diverse paths.

Comment [G174]: Synchronisation sometimes is used and other times coordination is used. I suggest using coordination rather synchronisation since synchronisation gives the impression that there is a time element involved.

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

- Clear switching command - the operator may request to clear previous administrative command to switch ~~over~~ (manual or force switch).

5.2. Fault Detection

Fault detection is a fundamental part of recovery and survivability. In all schemes except for some forms of 1+1 protection, the necessary actions for recovery of traffic delivery rely on discovering that there is some kind of fault.

Comment [M175]: Please explain this

Comment [G176]: Why?

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 of these may utilize control plane protocols (such as the routing protocols) to detect a failure of adjacency between neighboring nodes. In an MPLS-TP network, there is no certainty that a control plane will be present. Even if a control plane is present, it will be a GMPLS control plane [RFC3945] that makes a logical separation between control channels and data channels with the result that 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 5.4.1.

Faults may, however, be reported by lower layer. These generally show up as interface failures or link 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, and interface card failure may be reported to the MPLS-TP layer.

Comment [G177]: Or connectivity failure.

Such failures will only be reported after link level protection has been attempted (Section 4.6.1) and it is important that any lower layer recovery actions are coordinated with the MPLS-TP recovery actions (Section 4.6).

Comment [M178]: Defect reporting should not be conditioned by success/failure of recovery in the lowere layer.

Faults reported by lower layers are only visible at specific nodes within the MPLS-TP network (i.e. at the adjacent end-points of the MPLS-TP link). This only allows recovery to be performed locally.

In order that recovery can be performed by nodes that are not immediately local to the fault, the fault must be reported (Sections 5.4.3 and 5.5.4).

Comment [M179]: Please explain - if the node where the lower layer can report the defect to the MPLS layer is not at the boundary of an MPLS-TP recovery domain then the MPLS-TP layer ignores the defect report (or just inserts AIS). The MPLS-TP node at the recovery domain boundary will become aware of the failure as the result of a loss of CC/CV.

5.3. Fault Isolation

Comment [M180]: For the purposes of recovery faults are only detected at a recovery domain boundary and the egress node initiates recovery within the domain.

~~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:~~

~~o perform end-to-end recovery~~

~~o perform some other recovery as a speculative act.~~

~~Since speculative acts are not guaranteed to achieve the desired results and could be costly, and since end-to-end recovery is a costly option, it is important to be able to isolate the fault.~~

~~Fault isolation may be achieved by dividing the network into protection domains. End-to-end protection is thereby operated on an LSP segments depending on the domain in which the fault is discovered. This requires that the LSP can be monitored at the domain edges.~~

~~Alternatively, a proactive mechanism of fault isolation through OAM (Section 5.4.2) or through the control plane (Section 5.5.3) is required.~~

5.4. OAM Signaling

Comment [M181]: Delete this section and replace with a simple reference to the MPLS-TP OAM framework draft – explain that a MEP is coincident with the recovery domain boundary.

MPLS-TP provides 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).

These tools support proactive and on-demand fault management (for fault detection and fault localization) and for 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 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 [MPLS-TP-OAM-Framework], MPLS-TP OAM operates in the context of a Maintenance Entity which bounds the OAM responsibilities and represents the portion of a path between two points which is being monitored and maintained, and in which OAM messages are exchanged. [MPLS-TP-OAM-Framework] refers also to a Maintenance Entity Group (MEG), which is a collection of one or more MEs that belongs to the same transport path (e.g. P2MP transport path) and that are maintained and monitored as a group.

An ME includes two MEPs (Maintenance Group End Points) which reside at the boundaries of an ME, and a set of zero or more MIPS (Maintenance Group Intermediate Points) which reside within the Maintenance Entity along the path. A MEP is capable of initiating and terminating OAM messages, and as such can only located at the edges of a path where push and pop operations are supported. In order to be define an ME over a portion of path there is a need to support the MPLS-TP architectural element PST (Path Segment Tunnel). PST is an end-to-end LSP which corresponds in this context to the ME and makes use of the MPLS construct of hierarchical nested LSP which is defined in [RFC3031]. OAM messages can be initiated at the edge of the PST and sent to the peer edge of the PST, over G-ACH.

There is a need to configure the related PSTs and map between the LSP segment(s) being monitored and the PST. The 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 endpoints in the network and share similar constraints (such as requirements for QoS, terms of protection ,etc.).

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

MIP is capable of reacting to OAM messages.

Comment [G182]: Is the use of the term reacting correct?
MIP can only read but not write information

5.4.1. Fault Detection

Comment [M183]: Delete this section and replace with a simple reference to the MPLS-TP OAM framework draft – explain that a MEP is coincident with the recovery domain boundary.

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 misconnectivity (CV) allows a sink MEP can detect 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 CC&V (Continuity & Connectivity Verification) message every 3.33ms. In the absence of three consecutive CC&V messages, Loss of Continuity is declared and locally notified to the edge of the recovery domain 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 the OAM performance monitoring indicating degradation in the working entity may also be used as a trigger for protection switching. In the case of degradation, switching to the recovery entity is needed only if the recovery entity can guarantee better conditions. Degradation can be measured activating proactively the MPLS-TP OAM packet loss measurement or delay measurement.

- A MEP can get an indication from its sink MEP of a Remote Defect Indication and locally notify the endpoint of the recovery domain of fault condition to trigger the recovery action.

5.4.2. Fault Isolation

MPLS-TP provides OAM tools to isolate a fault and determining exactly where a fault has occurred. It is often the case the fault detection only takes place at key points in the network (such as at LSP end points, or MEPs). This means that the fault may be located anywhere within a segment of the LSP concerned. Finer granularity of information 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 isolate a fault.

The information may be locally notified to the endpoint of the recovery domain to allow ~~him~~ implementing optimal recovery action. This may be useful in case of re-calculation of a recovery path.

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

5.4.3. Fault Reporting

The endpoints of a recovery domain should be able to report to the ~~a network~~ management plane~~on~~ fault conditions detected in the recovery domain.

In addition, a node within a recovery domain detecting a fault condition should also be able to report the fault condition to the network management. The network management should be capable to correlate the fault reports and identify the source of the fault.

Comment [M184]: By definition recovery is between the edges of a recovery domain so how can additional knowledge be used?
It may be possible to invoke rerouting after the initial recovery action to improve resource utilization.

MPLS-TP OAM tools support a function where an intermediate node along a path can send an alarm report message to the MEP indicating of a fault condition in the server layer connecting it to its adjacent node. The purpose of this capability is to allow a MEP to suppress alarms that may be generated as a result of the failure condition in the server layer.

Comment [M185]: True, but how is this relevant to recovery?

5.4.4. Coordination of Recovery Actions

As described above, in some cases (such as in bidirectional protection switching, etc.) there is a need to coordinate the protection states between the edges of the recovery domain. [MPLS-TP-Linear-Protection] defines procedures and protocol messages and elements to support the PSC (Protection State Coordination) function.

Comment [M186]: What action is taken if the intended coordination does not occur?

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

The protocol also allow to detect mismatches between the configuration provisioned at the ends of the Protection Domain (such as timers, revertive/non-revertive behavior).

Comment [M187]: And reports miss match conditions?

5.5. 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 has been implemented and deployed in many networks, it is not surprising that it contains many features to support a high level of survivability function.

The signaling elements of the GMPLS control plane utilize extensions to the Resource Reservation Protocol (RSVP) as documented in a series of documents commencing with [RFC3471] and [RFC3473], but based on [RFC3209] and [RFC2205]. The architecture for GMPLS is provided in [RFC3945], and [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 isolation and reporting.

Clearly, the control plane techniques described here only apply where an MPLS-TP control plane is deployed and operated. All mandatory survivability features must be enabled even in the absence of the control plane, but where the control plane is present it may provide

Comment [G188]: Why do you still need control plane survivability features if there is no control plane?

alternative mechanisms that may be desirable by virtue of their ease of automation or richer feature-set.

5.5.1. Fault Detection

The control plane is not able to detect data plane faults. However, it does provide mechanisms to detect control plane faults and these ~~can be~~ can be used to deduce data plane faults where it is known 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 this is used exclusively. That means that there may be deployments where an in-band (or at least in-fiber) control channel is used. In this case, the failure of the control channel can be used to infer a failure of the data channel or at least 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 and, although these keep-alive mechanisms tend to operate at a relatively low frequency (order of seconds) it is still possible that the first indication of a control plane fault will be through the routing protocol.

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

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

Comment [M189]: Must be identical to the description of fault detection for protection. Note for high availability it is common to use both protection and restoration.

5.5.2. Testing for Faults

The control plane may be used to initiate and coordinate testing of links, LSP segments, or whole LSPs. This is important in some technologies where it is necessary to halt data transmission while testing, but 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, and 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

Comment [M190]: How does this apply to recovery? Why not simply use the already defined data plane OAM messages.

the other end of the link. This mechanism need not be disruptive to data flowing on the link.

Note that a link in 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 testing for faults. First, [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 provided in [GMPLS-OAM]. This protocol extension supports the configuration (including enabling and disabling) of OAM mechanisms for a specific LSP.

5.5.3. Fault Isolation

Comment [M191]: Not required for recovery action which by definition are only initiated at domain boundaries where a MEP is present to detect the fault.

Fault isolation is the process of determining exactly where a fault has occurred. It is often the case the fault detection only takes place at key points in the network (such as at LSP end points, or MEPs). This means that the fault may be located anywhere within a segment of the LSP concerned.

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

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

5.5.4. Fault Reporting

Comment [M192]: To which entity – fault reporting is normally to the OSS. This section appears to describe notification of status for distributed path computation. This is already described in other RFCs please provide the relevant references.

GMPLS signaling uses the Notify message to report faults. The Notify message can apply to a single LSP or can carry fault information for a set of LSPs 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 may be signaled as the LSP is set up. This allows the process to be made consistent with segment protection and

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 the PathErr and ResvErr messages.

[RFC4783] provides a mechanism to coordinate alarms and other event or fault information through GMPLS signaling. This mechanism is useful to understand the status of the resources used by an LSP and to help understand why an LSP is not functioning, but 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 withdraw the advertisement for a failed link, or may advertise that the link is about to be shut down gracefully. This mechanisms is, however, not normally considered to be fast enough to be used as a trigger for protection switching.

5.5.5. 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 remain to be defined for this operation.

A further piece of work in needed to allow control and configuration of reversion behavior for end-to-end and segment protection, and the coordination of timers' values.

5.5.6. Establishment of Protection and Restoration LSPs

It should not be forgotten that protection and recovery depend on the establishment of suitable LSPs. The management plane may be used to set up these LSPs, but the control plane may be used if it is present.

Several protocol extensions exist to make this process more simple:

- [RFC4872] provides features in support of end-to-end protection switching.

- [RFC4873] describes how to establish a single, segment protected LSP. Note that end-to-end protection is a sub case of segment protection and [RFC4872] can be used also to provide end-to-end

Comment [G193]: Why is this specific to this type of protection?

Comment [M194]: Is this for use by protection or restoration?

Comment [M195]: What is the purpose of stating this at this point in the draft?

protection.

- [RFC4874] allows one LSP to be signaled with a request that its path excludes specified resources (links, nodes, SRLGs). This allows a disjoint protection path to be requested, or a recovery path to be set up avoiding failed resources.

- Lastly, it should be noted that [RFC5298] provides an overview of the GMPLS techniques available to achieve protection in multi-domain environments.

6. Pseudowire Protection Considerations

Pseudowire is one of the clients of MPLS-TP. Pseudowires provide end-to-end connectivity over the MPLS-TP network and may be comprised of a single pseudowire segment, or multiple segments "stitched" together to provide end-to-end connectivity.

The pseudowire service may, itself, require a level of protection as part 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 in the PWE3 working group.

6.1. Utilizing Underlying MPLS-TP Recovery

MPLS-TP PWs are carried across the network inside MPLS-TP LSPs. Therefore, an obvious way to protect 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 speed of recovery necessary 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 on individual segments. That is, LSP recovery 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, the LSP protection of a PW is very much like the link-level protection offered to the MPLS-TP LSP layer by an underlying network layer (see Section 4.6).

Comment [M196]: But each segment is independent and protection may be active on more than one segment.

6.2. Recovery in the Pseudowire Layer

Recovery in the PW layer can be provided simply by running separate PWs either end-to-end or between S-PEs.

As with any recovery mechanism, it is important to coordinate between layers. This coordination is necessary to ensure that recovery mechanisms are only actioned 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), and to make 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 - compare with Section 4.6.2).

7. Manageability Considerations

TBD

8. Security Considerations

TBD

9. IANA Considerations

This informational document makes no requests for IANA action.

10. Acknowledgments

Thanks for useful comments and discussions to Italo Busi, David McWalter, Lou Berger and Yaacov Weingarten.

Formatted: German (Germany)

Formatted: German (Germany)

11. References

11.1. Normatve References

[RFC2205] Bradner, S., Ed., Zhang, L., Berson, S., Herzog, S., and J. Jamin, "Resource ReserVation Protocol &mdash 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., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions", RFC 3473, January 2003.

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

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

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

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

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

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

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

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

Formatted: Italian (Italy)

Formatted: Italian (Italy)

Formatted: Italian (Italy)

Formatted: Italian (Italy)

Formatted: Italian (Italy)

Formatted: Italian (Italy)

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

[MPLS-TP-FWK] Vigoureux, M., Ed., Ward, D., Ed., and Betts, "A Framework for MPLS in Transport Networks", MPLS-TP-FWK, Work in Progress.

[MPLS-TP-OAM] Buci, I., Ed. and B. Niven-Jenkins, Ed., "Requirements for OAM in MPLS Transport Networks", draft-ietf-mpls-tp-oam-requirements, Work in Progress.

[MPLS-TP-OAM-Framework] Buci, I., Ed. and B. Niven-Jenkins, Ed., "A Framework for MPLS in Transport Networks", draft-ietf-mpls-tp-oam-framework, Work in Progress.

11.2. Informative References

[RFC3031] Rosen, E., Viswanathan, A., and R. Callon, "Multiprotocol Label Switching Architecture", RFC 3386, November 2002.

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

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

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

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

[RFC4872] Lang, J., Rekhter, Y., and D. Papadimitriou, "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.

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

[MPLS-TP-Linear-Protection] Weingarten, Y., Bryant, S., Ed., Sprecher, N., Ed., Van Helvoort, H., Ed., and A. Fulignoli, "MPLS-TP Linear Protection", draft-weingarten-mpls-tp-linear-protection, Work in Progress.

[GMPLS-OAM] Takacs, A., Fedyk, D., and H. Jia, "OAM Configuration Framework and Requirements for GMPLS RSVP-TE", draft-ietf-ccamp-oam-configuration-fwk, Work in Progress.

[OAM-SOUP] Andersson, L., Betts, M., Van Helvoort, H., Bonica, R., and D. Romascanu, "MPLS-TP Linear Protection", draft-ietf-opsawg-mpls-tp-oam-def, Work in Progress.

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

Authors' Addresses

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

Email: nurit.sprecher@nsn.com

Adrian Farrel
Old Dog Consulting

Email: adrian@olddog.co.uk

Formatted: Spanish (Spain, International Sort)
Formatted: Spanish (Spain, International Sort)
Formatted: Spanish (Spain, International Sort)
Formatted: Spanish (Spain, International Sort)
Formatted: Spanish (Spain, International Sort)