Technical Specification
MEF 2

Requirements and Framework for Ethernet Service Protection in Metro Ethernet Networks

Feb 8, 2004

Disclaimer

The information in this publication is freely available for reproduction and use by any recipient and is believed to be accurate as of its publication date. Such information is subject to change without notice and the Metro Ethernet Forum (MEF) is not responsible for any errors. The MEF does not assume responsibility to update or correct any information in this publication. No representation or warranty, expressed or implied, is made by the MEF concerning the completeness, accuracy, or applicability of any information contained herein and no liability of any kind shall be assumed by the MEF as a result of reliance upon such information.

The information contained herein is intended to be used without modification by the recipient or user of this document. The MEF is not responsible or liable for any modifications to this document made by any other party.

The receipt or any use of this document or its contents does not in any way create, by implication or otherwise:

(a) any express or implied license or right to or under any patent, copyright, trademark or trade secret rights held or claimed by any MEF member company which are or may be associated with the ideas, techniques, concepts or expressions contained herein; nor

(b) any warranty or representation that any MEF member companies will announce any product(s) and/or service(s) related thereto, or if such announcements are made, that such announced product(s) and/or service(s) embody any or all of the ideas, technologies, or concepts contained herein; nor

(c) any form of relationship between any MEF member companies and the recipient or user of this document.

Implementation or use of specific Metro Ethernet standards or recommendations and MEF specifications will be voluntary, and no company shall be obliged to implement them by virtue of participation in the Metro Ethernet Forum. The MEF is a non-profit international organization accelerating industry cooperation on Metro Ethernet technology. The MEF does not, expressly or otherwise, endorse or promote any specific products or services.

Table of Contents

1 Abstract 4
2 Terminology 4
3 Scope 7
4 Compliance Levels 7
5 Introduction 7
6 Protection Terminology 10
6.1 Protection Types 10
6.2 Failure Types 10
6.3 Resource Selection 11
6.4 Event Timing 11
6.5 Other Terms 13
7 Discussion of Terminology 13
7.1 Timing Issues 13
7.2 SLS Commitments. 14
8 Protection Reference Model 14
8.1 Transport 15
8.2 Topology 16
8.3 MEF Protection Mechanism 17
8.4 Link Protection based on Link Aggregation 23
8.5 Application Protection Constraint Policy (APCP) 24
9 Requirements for Ethernet Services protection mechanisms 24
9.1 Service-Related Requirements 24
9.2 Network Related Requirements 25
10 Framework for Protection in the Metro Ethernet 28
10.1 Introduction 28
10.2 MEF Protection Schemes 28
11 Requirements summary 31
12 Appendix A: Transport Protection 33
12.1 General 33
12.2 Layered protection characteristics 34
12.3 Potential problems of protection interworking 36
12.4 Methods for internetworking between layers 37
13 Appendix B Transport Indications 37
13.1 Optical transmission HW indications 38
13.2 Ethernet HW indications 38
13.3 Ethernet-specific counters based decisions 38
13.4 SONET/SDH indications 38
13.5 RPR indications 38
14 Appendix C (informative): Restoration Time Requirements derived from Customer Ethernet Control Protocols 39
14.1 Spanning Tree Protocol and Rapid Spanning Tree Protocol 39
14.2 Generic Attribute Registration Protocol 39
14.3 Link Aggregation Control Protocol 40

15 References 40

List of Figures

Figure 1: Illustration of event timing 13
Figure 2: The PRM model (two layers are shown, from a stack of two or more) 15
Figure 3: ALNP 18
Figure 4: EEPP 19
Figure 5: Split Horizon bridging with full mesh connectivity 21
Figure 6: Link-redundancy 23
Figure 7: Link-aggregation 23
Figure 8: Failure that can be restored by SONET/SD or RPR 35
Figure 9: Failure cannot be repaired by SONET/SDH (BLSR, UPSR) or RPR; it can be repaired by MPLS 35
Figure 10: Failure can be restored by the ETH layer 35
Figure 11: Simple network with protection at different layers 36
Figure 12: Protection by EEPP performed, but not required 36
Figure 13: Final result for revertive switching 37

1 Abstract

This document provides requirements, a model, and a framework for discussing protection in Metro Ethernet Networks.

2 Terminology

Access Link	A link that represents connectivity to External Reference Points of the MEN
ADM	Add Drop Multiplexer
ALNP	Aggregated Line and Node Protection
APCP	Application Protection Constraint Policy
APS	Automatic Protection Switch
BER	Bit Error Rate
BLSR	Bi-directional Line Switching Redundancy
BPDU	Bridge Protocol Data Unit
CE	Customer Equipment
CES	Circuit Emulation Service
CIR	Committed Information Rate
CRC	Cyclic Redundancy Check
CSPF	Constriant-based Shortest Path First

DCE	Data Circuit-terminating Equipment
DSL	Digital Subscriber Line
ECF	Ethernet Connection Function
EEPP	End-to-End Path Protection
EFM	Ethernet First Mile
EIR	Excess Information Rate
E-Line	Ethernet Line Service
E-LAN	Ethernet LAN Service
EoS	Ethernet over Sonet
ETH	Ethernet Services Layer
ETH-trail	An ETH-trail is an "ETH-layer entity" responsible for the transfer of information from the input of a trail termination source to the output of a trail termination sink.
EVC	Ethernet Virtual Connection
GARP	Generic Attribute Registration Protocol
GRE	Generic Routing Encapsulation
IETF	Internet Engineering Task Force
IGP	Interior Gateway Protocol
ITU	International Telecommunication Union
LAG	Link Aggregation Group
LAN	Local Area network
LACP	Link Aggregation Control Protocol
LAG	Link Aggregation Group
Link	An ETH link or TRAN link
LOF	Loss of Frame
LOS	Loss of Signal
LSP	Label Switched Path
LSR	Label Switched Router
MAC	Media Access Control
Mean time to restore	The mean time from when a service is unavailable to the time it becomes available again
MEF	Metro Ethernet Forum

MEN	Metro Ethernet Network
MPLS	Multi-Protocol Label Switching
NE	Network Element
Node	A Provider owned network element
OAM	Operations, Administration and Maintenance
Path	A succession of interconnected links at a specific (ETH or TRANS) layer
PE	Provider Edge
PRM	Protection Reference Model
Protection merge point	A point in which the protection path traffic is either merged back onto the working path or passed on to the higher layer protocols (used in [3], called 'tail-end switch' in SONET/SDH).
QoS	Quality of Service
RPR	Resilient Packet Ring
RSTP	Rapid Spanning Tree Protocol
SDH	Synchronous Digital Hierarchy
Segment	A connected subset of the trail
SLA	Service Level Agreement
SLS	Service Level Specification
SONET	Synchronous Optical Network
SRLG	Shared Risk Link Group
STP	Spanning Tree Protocol
Subscriber	The organization purchasing and/or using Ethernet Services. Alternate term: Customer
TCF	Transport Connection Function
TCP	Transmission Control Protocol
TDM	Time Division Multiplexing
TRAN	Transport Services Layer
Transport	A specific TRANS layer technology
TRAN-trail	A TRAN-trail (see ITU-T Recommendation G.805) is a "transport entity" responsible for the transfer of information from the input of a trail termination source to the output of a trail termination sink.
TTF	Trail Termination Function

UNI	User to Network Interface
UNI N	A compound functional element used to represent all of the functional elements required to connect a MEN to a MEN subscriber implementing a UNI C.
UNI C	A compound functional element used to represent all of the functional elements required to connect a MEN subscriber a MEN implementing a UNI N.
User Network Interface	The demarcation point between the responsibility of the Service Provider (UNI N) and the responsibility of the Subscriber (UNI C).
WTR	Wait to Restore

3 Scope

The scope of this document is to provide requirements to be satisfied by the protection and restoration mechanisms for Ethernet services in Metro Ethernet Networks and a model and framework for discussing protection mechanisms for Ethernet services-enabled architectures in Metro Networks. The document discusses requirements from the network according to the service it provides regardless of the specific implementation, and provides the model framework for mechanisms that provide protection to Ethernet Services in MENs according to these requirements. It is the objective of the document to provide requirements, model, and framework that are as much as possible independent of a given transport.

Some customers desire reliability and redundancy in the attachment of the CE to the network. This usually requires dual homing to the provider network as well as requirements on the CE. The different CE-attachment redundancy mechanisms are not in the scope of this document. In other words, this document does not apply to CE. In the case of subscriber access connections the requirements, model, and framework described in this document apply until the UNI or edge of UNI N.

4 Compliance Levels

The key words "**MUST**", "**MUST NOT**", "**REQUIRED**", "**SHALL**", "**SHALL NOT**", "**SHOULD**", "**SHOULD NOT**", "**RECOMMENDED**", "**MAY**", and "**OPTIONAL**" in this document are to be interpreted as described in RFC 2119. All key words must be use upper case, bold text.

5 Introduction

Protection in Metro Ethernet Networks (MEN) can encompass many ideas. Basically, it is a self-healing property of the network that allows it to continue to function with minimal or no impact to the network users upon disruption, outages or degradation of facilities or equipment in the MEN. Naturally there is a limit to how much the network can be disrupted while maintaining services, but the emphasis is not on this limit, but rather on the ability to protect against moderate failures.

Network protection can be viewed in two ways:

- From the viewpoint of the user of the MEN services [1], the actual methods and mechanisms are of minor concern and it is the availability and quality of the services that are of interest. These can be described in a Service Level Specification (SLS), a technical description of the service provided, which is part of the Service Level Agreement (SLA) between customer and provider.
- The other viewpoint is that of the network provider. The provider is tasked with translating the SLSs of all the customers (and future customers) into requirements on the network design and function. We do not study this translation here; it is an area of differentiation and specialization for the provider and depends on the policies that the provider will use for protection. What we do study is the mechanisms that can be used to provide protection.

Any protection scheme has three clear components:

- Detection: refers to the ability to determine network impairments.
- Policy: defines is what should be done when impairment is detected.
- Restoration is the component that acts to "fix" the impairment; it may not be a total restoration of all services and depends on the nature of the impairment and the policy.

We focus on the detection and restoration mechanisms and leave the choice of policy to the providers. However, the policy itself cannot be ignored and is based on the services supported.

Detection and restoration can be done in many different ways in the MEN. The techniques available depend on the nature of the equipment in the network.

The requirements have basis in the interpretation of Service Level Specifications for Ethernet services (such as availability, mean time to restore, mean time between failure, etc.) in terms of network protection requirements (such as connectivity restoration time, SLS restoration time, protection resource allocation, etc.). This means that the protection offered by the network is directly related to the services supplied to the user and the requirements derived from the need to protect the services provided to the user.

In most cases, an EVC implementing an Ethernet service traverses different transports and therefore the end-to-end protection may involve different mechanisms. For example, many transports may be involved: Ethernet, Ethernet over DSL, Ethernet over SONET/SDH, MPLS [5], [3] and data link layer switching as Ethernet [11]. In the case of Ethernet protection, technologies such as RSTP [802.1w] or Link Aggregation [11] may be used to provide protection at the ETH layer.

An Ethernet Line service EVC is built of a single ETH-trail, while an Ethernet LAN service EVC is built of a number of ETH-trails.

The details of the protection mechanisms will therefore vary throughout the network and it is in the scope of the MEF to describe how each portion of the network with its specific transport and

topology can be protected and how the different protection mechanisms present in the network will interwork.

However the scope of the requirements presented in this document is more limited. The document only discusses requirements from the network according to the service it provides regardless of the specific implementation. It is the objective of the document to provide requirements that are as much as possible independent of a given transport.

The protection requirements section provides requirements with two distinct goals, both of which are covered throughout the document. Protection requirements are specified for Service Level Specifications (such as protection switching time) and can be measurable parameters, which can be specified in SLSs. Other protection requirements are specified for providers of the Service (such as Protection Control Requirements specifying protection configuration), and are not directly reflected in a Service Level Specification, but are required from the provider. Examples for such requirements are those that relate to control, manageability, and scalability of a protection scheme.

The following topics are examples of those discussed in the requirements section:

- Protection switching times;
- Failure detection requirements;
- Protection resource allocation requirements;
- Topology requirements;
- Failure notification requirements;
- Restoration and revertiveness requirements;
- Transparency for end-user;
- Security requirements: e.g., separation between LAN & MAN protection mechanisms.

Observe that if all EVCs passing through a specific connected part of the network are known to have similar protection requirements, it is sufficient for this part of the network to comply with the specific requirements that are needed by the EVCs of services passing through it. An example is the “last-mile”: protection requirements are directly related to the customers needs.

The framework defined in this document deals with models and mechanisms specific to the Metro Ethernet. We can make use of any existing mechanisms for protection of transport, and that upper-layer protection mechanisms can sit on top of lower-layer protection mechanisms to provide a unified protection approach. This is much clearer once we look more closely at a model for protection, presented in section 8. The model allows protection mechanisms to be enabled as part of each layer (ETH layer or TRAN layer) in the network. Sections 6 and 7 discuss the terminology used in this document. The remainder of the paper focuses on setting the requirements and on a framework for the protection mechanisms. Discussion of the transport layer and interworking between layers is presented in an appendix as well as a discussion of the requirement imposed by customer Ethernet control protocols.

6 Protection Terminology

This section defines the precise terminology that will be used in all MEF protection documents.

6.1 Protection Types

A network can offer protection by providing alternative resources to be used when the working resource fails. There is specific terminology for the number and arrangement of such resources.

6.1.1 1+1

The Protection Type 1+1 uses the protection resources at all times for sending a replica of the traffic. The protection merge point, where both copies are expected to arrive, decides which of the two copies to select for forwarding. The decision can be to switch from one resource to the other due to an event like resource up/down etc. or can be on a per frame/cell basis, the selection decision is performed according to parameters defined below (e.g. revertive, non-revertive, manual, etc.).

6.1.2 m:n

The m:n Protection Type provides protection for n working resources using m protection resources. The protection resources are only used at the time of the failure. The protection resources are not dedicated for the protection of the working resources, meaning that when a protection resource is not used for forwarding traffic instead of a failed working resource, it may be used for forwarding other traffic. The following subsections define the important special cases of m:n protection.

There are two variants of m:n protection type, one in which a protection resource can be used concurrently for forwarding the traffic of a number of working resources, in case a few of them fail at the same time. The other variant is when the protection resource is able to forward the traffic of a single working resource at a time.

6.1.2.1 1:1

The 1:1 Protection Type provides a protection resource for a single working resource.

6.1.2.2 n:1

The n:1 Protection Type provides protection for 1 working resource using n protection resources.

6.1.2.3 1:n

The 1:n Protection Type provides protection for n working resources using 1 protection resource. In this protection type, the protection resource is shared for protection purposes by the n working resources.

6.2 Failure Types

Failures may occur in network nodes or on the links between nodes.

6.2.1 Fail condition (Hard Link Failure)

Fail condition is a status of a resource in which it is unable to transfer traffic (e.g. Loss of Signal, etc.).

6.2.2 Degrade condition (Soft Link Failure)

Degrade Condition is a status of a resource in which traffic transfer might be continuing, but certain measured errors (e.g., Bit Error Rate, etc.) have reached a pre-determined threshold.

6.2.3 Node Failure

A Node Failure is an event that occurs when a node is unable to transfer traffic between the links that terminate at it.

6.3 Resource Selection

6.3.1 Revertive Mode

The protection is in revertive mode if, after a resource failure and its subsequent repair, the network automatically reverts to using this initial resource. The protection is in non-revertive mode otherwise. Automatic reversion may include a reversion timer (i.e., the Wait To Restore), which delays the time of reversion after the repair.

6.3.2 Manual Switch

A Manual Switch is when the network operator switches the network to use the protection resources instead of the working, or vice-versa. By definition, a Manual Switch will not progress to failed resources. Manual switch may occur at any time according to the network operator will, unless the target resource is in failure condition.

6.3.3 Forced Switch

A Forced Switch is when the network operator forces the network to use the protection resources instead of the working resources, or vice-versa, regardless of the state of the resources.

6.3.4 Lockout

A lockout command on a resource makes the resource not available for protection of other resources.

6.4 Event Timing

The terminology distinguishes events, which occur at particular instants (points in time), and times, which are the time durations between events.

6.4.1 Impairment Instant

The Impairment Instant is the point in time that the failure event occurs.

6.4.2 Fault Detection Instant

The Fault Detection Instant is the point in time at which the failure is detected and declared. The Fault Detection Instant may be different for different network elements.

6.4.3 Hold-off Instant

It may be desirable to delay taking any action after the Fault Detection Instant. The Hold-off Instant is the instant at the end of this delay period, if there is one. Otherwise it is the same as the Fault Detection Instant. Hold-off instant is useful when two or more layers of the same network provide protection. In such a case, the hold-off instant in an upper layer gives an opportunity to the protection in a lower-layer to take place before the upper layer protection acts.

6.4.4 Connectivity Restoration Instant

The Connectivity Restoration Instant is the first point in time after impairment that user traffic can begin to be transferred end-to-end. At the connectivity restoration instant, services affected by the failure already deliver user-traffic end-to-end but may not do that in the performance required by the SLS.

In case the impairment caused only degradation in performance to the point of losing SLS compliance, the connectivity restoration instant is defined to be identical to the impairment instant.

6.4.5 SLS Restoration Instant

The SLS Restoration Instant is the first point in time after impairment that user traffic can begin to be transferred end-to-end with the original performance guarantees.

6.4.6 Reversion Instant

In revertive mode, the Reversion Instant is the point in time at which the original resources are again used. This point in time MAY be the same as the SLS Restoration Instant.

6.4.7 Detection Time

The Detection Time is the difference between the Fault Detection Instant and the Impairment Instant.

6.4.8 Hold-off Time

The Hold-off time is the difference between the Hold-off Instant and the Fault Detection Instant. The Hold-off time may be zero.

6.4.9 Connectivity Restoration Time

The Connectivity Restoration Time is the difference between the Connectivity Restoration Instant and the Impairment Instant.

6.4.10 SLS Restoration Time

The SLS Restoration Time is the difference between the SLS Restoration Instant and the Impairment Instant.

6.4.11 Reversion (wait to restore (WTR)) Time

In revertive mode, the Reversion Time is the difference between the repair instant of the original resource and the Reversion Instant.

6.4.12 Timing Relationships

Figure 1, below, shows event instants and times, as defined above, on a timeline. Some times shown may actually be identical. Other times besides those defined above may be of interest.

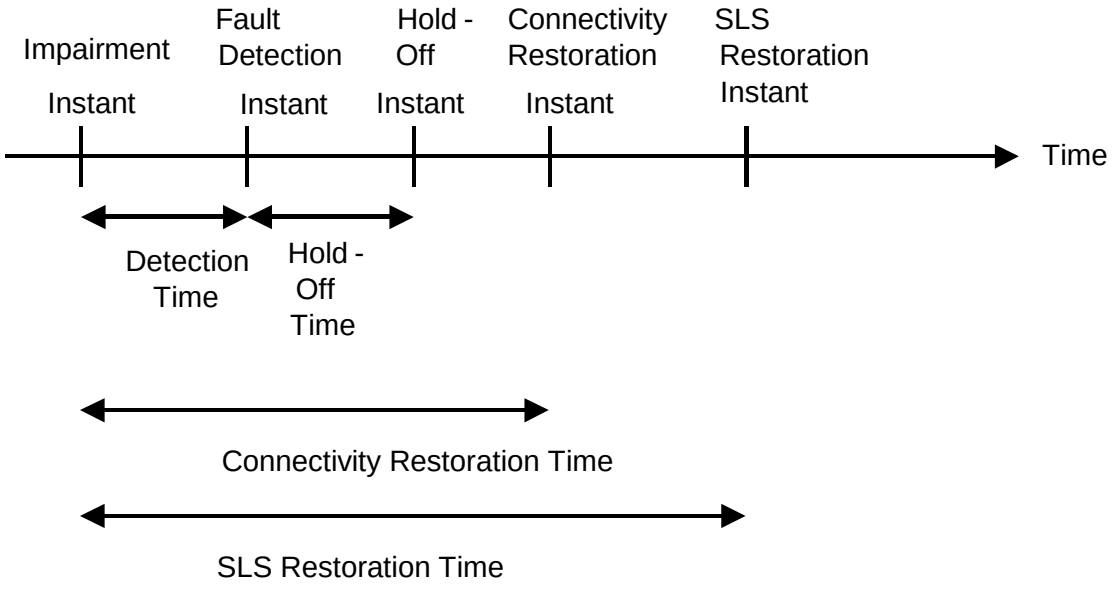

Figure 1: Illustration of event timing

6.5 Other Terms

6.5.1 Shared-Risk Link Group (SRLG)

A Shared Risk Link Group (SRLG) is a group of links that share a resource whose failure affects all the links in the group. For example, a bundle of fibers connecting two sites is an SRLG.

7 Discussion of Terminology

7.1 Timing Issues

Different applications and different users require different restoration times. In some provider network designs, a faster restoration time may require more network resources. For example, one technique of providing fast restoration is by creating one or more protection paths per working

path that needs to be protected, including the provisioning of bandwidth for the protection path. For this reason, it is beneficial to define a number of different restoration times that can be provided for different resources.

7.2 SLS Commitments.

Protection paths can either maintain or reduce bandwidth, and can alter other SLS characteristics. For example, a path that is assigned with a certain amount of CIR can be assigned with EIR instead when going through the protection path. In this way, the protection path requires fewer resources to be provisioned.

In [3] and [5], two different types of protection times are mentioned, ”recovery time” and “full restoration time.” These correspond to the Connectivity Restoration Time and the SLS Restoration Time defined above. The differentiation in the terms is based on the SLS provided on the protection path, as well as the time it takes to provide the SLS provided by the original working path. In terms of providing various SLS levels on the protection path, it can be beneficial to provide a 2-stage protection mechanism where the first stage protection switching occurs (rapidly) on to the “limited protection path” which is a protection path with reduced SLS commitment. The “equivalent protection path” which is a protection path with full SLS commitment can then be installed and the traffic is switched from the first-stage protection path. Note that, as also mentioned in [3], such “full restoration time” (SLS restoration time) may or may not be different from the “recovery time” (connectivity restoration time) depending on whether limited or equivalent protection path is used as the first-stage protection switching.

The protection type 1+1 uses the resources in the protection path at all times for sending a replica of the traffic. The protection merge point decides which of the two copies to forward. On the other hand, 1:1, 1:n, n:1, and m:n protection use only one path at a time, and therefore have the advantage that the protection-provisioned bandwidth can be used for other purposes when there is no failure.

8 Protection Reference Model

To deliver protection to Ethernet services implemented over Metro Ethernet Networks (MEN), a reference model has been created to allow description of a unified protection structure. The Protection Reference Model (PRM) allows a consistent description of protection capabilities applied on these services across various transmission technologies, network topologies and policies, thereby enabling the description of protection of services in the ETH layer (Ethernet Line Service, Ethernet LAN service, etc.).

The following PRM section highlights the main functional elements of protection in a MEN. The elements are described in following sections in a “bottom-up” approach. The model shows a single layer in the architecture, which can be a transport-layer or the Eth layer. The Trans layer is built of layered trails, so there can actually be layering of transports, and protection can be provided at each of these layers (for example Ethernet over MPLS over SONET). This is shown in the figure below: each layer may contain protection capabilities, and may run above a lower

layer that might contain protection capabilities as well. The entire protection scheme is controlled according to the application protection constraint policy.

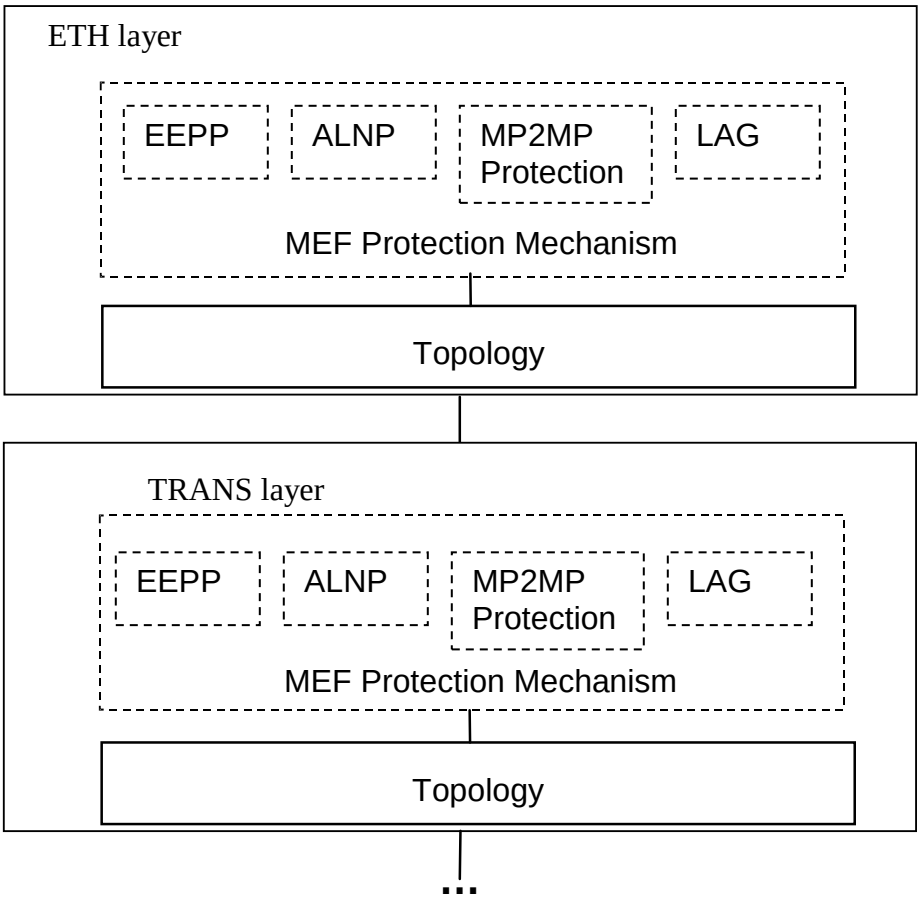

Figure 2: The PRM model (two layers are shown, from a stack of two or more)

8.1 Transport

The purpose of the Transport Layer is to provide transfer of data between MEN elements. Many transports provide error checking mechanism and data flow controls. The Transport Layer leverages any native mechanisms that the transport technology provides to gain protection from faults. The type of protection available may be local (e.g., a given node or link) or end-to-end (e.g., a set of nodes and links traversed by a "virtual" link) depending upon the technology used.

The scenarios of the MEF protection architecture can be divided into two categories:

1) The service is carried natively over the transport layer and protection is achieved at the transport layer. An example is carrying Ethernet traffic in Ethernet over SONET (EoS), where the protection is done at the SONET layer.

2) The protection is done above a transport layer. Here there are two sub-types:

    a. A transport layer is not capable of providing protection, or its protection capability is ignored by the protection mechanism of the upper layer. An example is Ethernet transport with the protection performed at the ETH layer or in an MPLS layer above it.

    b. A transport layer and the protection mechanism of the upper layer are working in conjunction to bring the required protection SLS. An example is where the ETH layer containing a protection mechanism is implemented over an interconnected succession of SONET transport networks with 1+1 capability. The SONET 1+1 capability repairs local failures on certain links, while ETH layer protection is used where SONET protection is not applicable or as an additional end-to-end protection method.

The second case is described in more detail in Appendix A

The ability of protection mechanisms to be independent of the transport technologies allows metro networks to be deployed utilizing various transmission technologies, interconnected to create a heterogeneous network infrastructure. Section 9 specifies that consistent protection-related Service Level Specification (SLS) SHOULD be delivered end-to-end in a metro network (or across national or regional levels) for higher-level services to be meaningful. Protection mechanisms can span across various transmission technologies (transports) regardless of whether each of these transports can deliver native protection capabilities. As each individual subnetwork of transport is utilized in a MEN, protection mechanisms could be requested from these transports to match an end-to-end protection SLS. If a transport does not have the ability to offer such services, then protection capabilities are performed at a higher or a lower layer, to ensure end-to-end protection SLS.

8.2 Topology

Protection requires the topology to be such that does not hinder an end-to-end protection SLS from being supported: Section 9 specifies that the protection scheme **SHOULD** support different topologies, although a specific mechanism may be limited to few topologies. (A sparse topology with no redundancy, i.e. a topology in which a resource does not have a path excluding itself connecting each pair of its neighboring resources, cannot offer protection, but any sufficiently rich topology is sufficient for end-to-end protection.) However, this requirement does not mean that the topology at a specific layer should be one that allows protection, as long as this part of the network is protected either at a lower layer or at a higher layer.

Depending on the specific technology, topology discovery may also be important to ensure that nodes (or a management utility) understand how to support the required protection. There can be

many ways of delivering topology discovery including an Interior Gateway Protocol (IGP) with topology extensions. The topology is different at each layer of the MEN, as the internal topology of the lower layer is not visible to the upper layer and vice versa.

The topology may look different when looking at different layers of the network. At the ETH-layer, the network is built of ECFs, interconnected by ETH-links, where each ETH-link is implemented as a trail in the layer below. Different example topologies at the ETH layer are:

- ECFs on the edges only.
- ECFs at edges and in the core (e.g. grooming of Ethernet service frames for efficiency improvement, where EVCs are supported using multiple transport layers).

Each TRAN-layer subnetwork over which the ETH layer is implemented has its own topology, built of TCFs interconnected by TRAN-links.

Protection can be provided in a specific layer if the topology at that layer contains enough redundancy. A service can be protected even if the topology at a specific layer does not provide enough redundancy, as long as the protection at other layers creates an end-to-end protection for the service at the ETH-layer.

The rest of the document contains a generic discussion from the point of view that the mechanisms described possibly apply to a few layers and technologies. For this reason, the document uses the terms links, nodes and Network Elements, where:

- Network Element (NE, node) refers to a device containing an ECF or an TCF depending on the layer.
- Link refers to an ETH-link or a TRAN-link depending on the layer.

8.3 MEF Protection Mechanism

The following styles of network protection mechanisms are currently under consideration:

1. Aggregated Line and Node Protection (ALNP) service
2. End-to-End Path Protection (EEPP) service
3. MP2MP protection service
4. Link Protection based on Link Aggregation

The protection services can be layered one on top of the other in any combination. For example, the ALNP can protect the network facilities while EEPP provides an additional protection at the path level.

EEPP supports 1+1, 1:1, and 1:n protection mechanisms, ALNP supports 1:1 as well as 1:n facility protection.

8.3.1 Aggregated Line and Node Protection (ALNP)

ALNP provides protection against local link and nodal failure by using local path detour mechanisms. In this case, local “backup” or “detour” paths are created along the primary path, that bypass the immediate downstream network element NE or the logical link and immediately merge back on to the primary path. The detour path may provide 1:n protection or 1:1 protection of the primary paths in the network. The backup paths are either explicitly provisioned, as

described as an option in [3] or are implicit in the technology, as in SONET/SDH ULSR/BLSR [6].

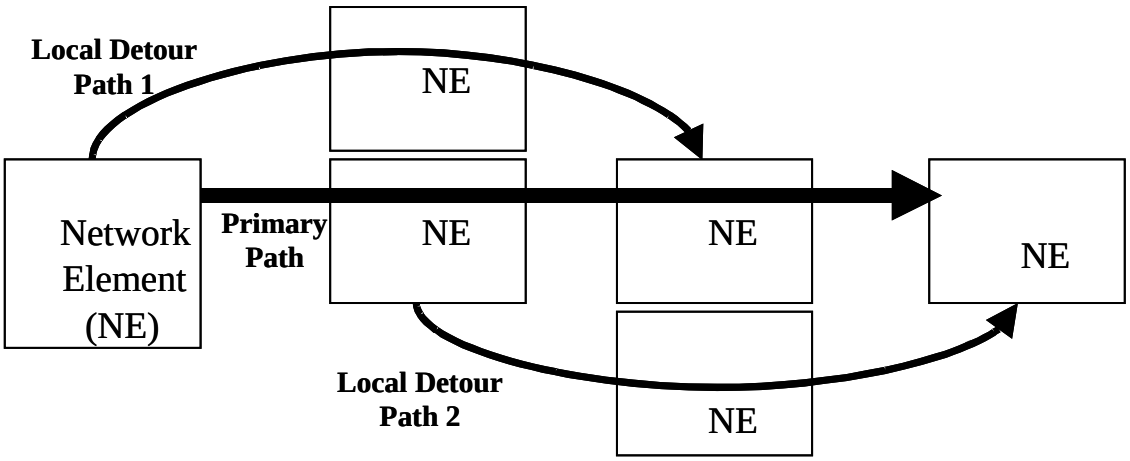

Figure 3: ALNP

Protection with short restoration times is possible in many cases with ALNP because many failure events can be instantaneously detected without long feedback loops end-to-end. The restoration time actually depends on the local failure detection time, which may differ in different scenarios. As each failure is detected (at each link or node) ALNP protects many end-to-end paths (with similar end-to-end protection SLSs) in a single restoration invocation. If a lower layer transport subnetwork has the ability to deliver services that are similar to those that ALNP provides at an upper layer, then the native protection mechanism of the transport subnetwork can be used and ALNP can be bypassed.

If a transport subnetwork in a layer below the layer at which ALNP operates does not support native protection capabilities to support a specified SLS, then it is the responsibility of the Aggregated Line and Node Protection (ALNP) mechanism to deliver the protection required according to the specified SLS.

Bi-directional Line Switching Redundancy (BLSR) capabilities of SONET and SDH [6], and MPLS local repair, described as an option in [3] are examples of ALNP derivatives in specific transports. ALNP may deliver a 1:n protection capability with a sub50ms restoration time and other default parameters. (Other restoration times could also be supported and invoked depending on the protection SLS specified and on failure detection capabilities.) As discussed in Section 9 the protection SLS that ALNP SHOULD deliver will be dependent on the protection desired for the service or services it protects. ALNP provides the ability to aggregate many end-to-end paths in a hop-by-hop and node-by-node manner. At any time both the ALNP and other protection mechanisms in transport layers below the layer at which ALNP executes could offer similar protection capabilities. Interoperability is achieved in this case by configuration of the

hold-off time of the ALNP mechanism such that the lower layer protection mechanism converges before the ALNP mechanism at the upper layer decides whether to take action.

To protect each link and node using ALNP, the mechanism for generating ALNP protection paths is preferably done using an automated scheme or is implicit in the transport technology. A possible mechanism for automatic creation of protection paths is to allow the specification of the protection parameters desired as part of the trail creation. Upon detection of the first request (for a given protection SLS for a specific trail), or earlier (e.g. at network setup), protection paths with certain protection parameters are created for each given transport subnetwork.

8.3.2 End-to-End Path Protection (EEPP)

End-to-end path protection (EEPP) is the ability to provide a redundant end-to-end path for the primary path. This mechanism can be used to augment ALNP. A variation of this method can be used to protect partial segments of the end-to-end path within the same layer if such capability is supported by the protection mechanism at the specific layer.

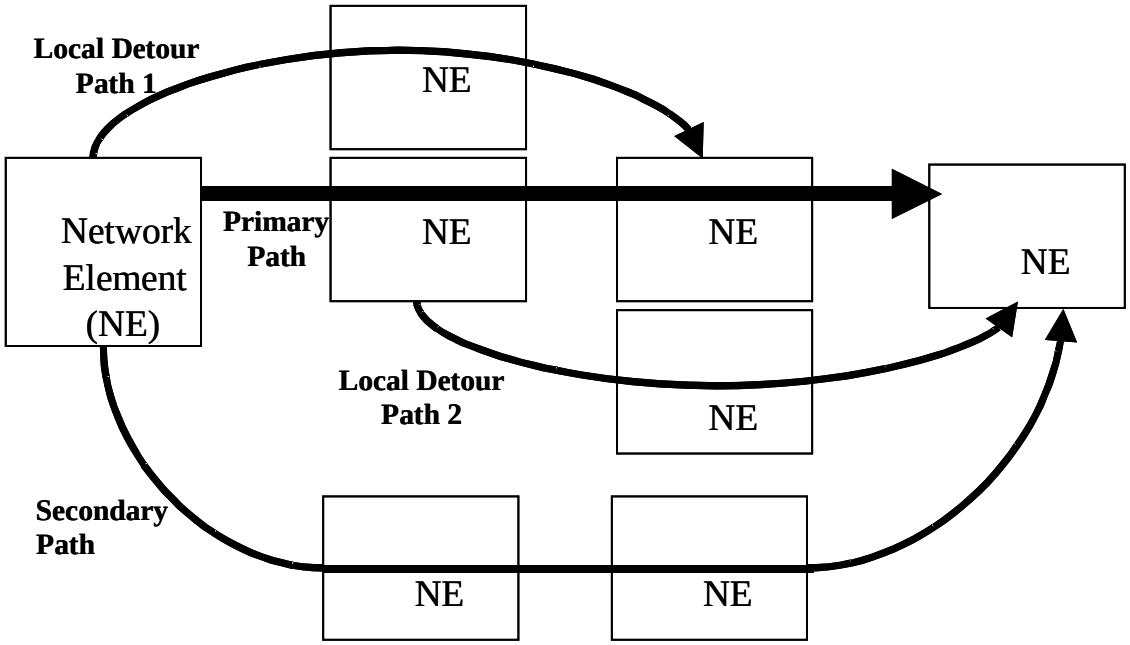

Figure 4: EEPP

The diagram above illustrates the use of a secondary path for EEPP as well as detour paths for ALNP.

In an EEPP scenario, a path is created from a source node to a destination node. Alternative or secondary paths are then created with different routing segments that protect the primary path. The number of redundant paths needed is policy-defined and has implementational limits (each of these redundant paths could consume network resources such as bandwidth, CPU cycles and memory). The intelligence for computing redundant paths (that are not part of the primary path resources) can be done with an online constraint mechanism (e.g., CSPF) or offline traffic

engineering tools. Each of these redundant paths is end-to-end in nature and therefore provides redundancy in a different manner than ALNP. The EEPP handles protection on segments of the global path, and in some cases could be provisioned end-to-end, and can provide redundancy when a transport segment along the path cannot provide protection of any kind (this includes ALNP or native transport protection). The EEPP could also be used when ALNP protection is available at each transport subnetwork but further redundancy is desired for path diversification. The restoration time of EEPP can be much longer than for ALNP and is dependent on the protection-type that is used. There are a few types of protection that can be used:

1+1– This type of configuration requires that the redundant paths are fully established and active. All data sent on the primary path is also copied to the redundant path(s). The receiver (which is the node at which the two paths merge) coordinates and decides which of all the available paths (primary and secondary) is used at each point in time. This decision can be performed on a per path basis according to OA&M for example, or on a per packet basis, where each packet is taken from a path from which it was received, in which case a sequence-number field can be added to the packets so that the receiver can correlate between the two packet streams. This type of redundancy can achieve very fast restoration times (milliseconds) since the receiver decides if the primary path has failed based on alarm indication or performance information derived from the primary path. However this type of redundancy will consume double the bandwidth and hardware resources (CPU, fabric, memory, etc.) as it is always active and passing data.

1:1 Cold Standby - This type of configuration requires that the redundant paths have their routing information calculated ahead of time, but the redundant paths are not established until failure of the primary path; the source node establishes the redundant path only when failure has occurred on the primary path, resulting in long restoration times.

1:1 Hot Standby - This type of configuration requires that the redundant paths have their routing information calculated ahead of time and established during the service activation time of the primary path: the redundant path(s) are kept active waiting for the primary path to fail. The chief determination of the time to repair a failure is the detection time since the switch over to any redundant path(s) can occur very quickly. The draw back to this type of redundancy is that the redundant paths consume network resources even though they are not passing data. Based on protection policy however, the set-up of the redundant paths may be made with fewer resources in order to give fast restoration for part of the traffic immediately. Cold standby can be invoked later to restore full traffic BW.

Shared Redundancy – Since a single failure in the network may only affect a subset of the primary paths, there is an opportunity to share same protection resource among multiple primary paths. There are many schemes that achieve sharing of the protection resources by exploiting this fact. 1:n, ring, and shared mesh protection are some of the well-known sharing mechanisms.

8.3.3 MP2MP protection

The E-LAN service is a multipoint-to-multipoint service that requires connectivity between all its UNIs. Depending on the implementation of the E-LAN service, the protection schemes above may not be sufficient for protecting it. The reason is that the implementation of an E-LAN may

be involved with one or more ECFs, which are interconnected by a number of ETH-trails. A failure of such an ECF is not covered by EEPP and ALNP as described above. The implementation of an E-LAN service may include implementation of multipoint-to-multipoint connectivity at the TRANS-layer as well. Three methods are typically used for multipoint-to-multipoint protection of Ethernet service or transport:

- Split Horizon bridging with full mesh connectivity.
- Spanning Tree or Rapid Spanning Tree.
- Link Redundancy.

With Split-Horizon bridging a full mesh topology is created between the TTF (Trail Termination Function) entities (Each is ECF or an TCF depending on the layer under discussion) creating the protected domain. Each trail in the full-mesh of trails is a point-to-point trail and may contain nodes (ECFs or TCFs) in the same layer or in a lower layer.

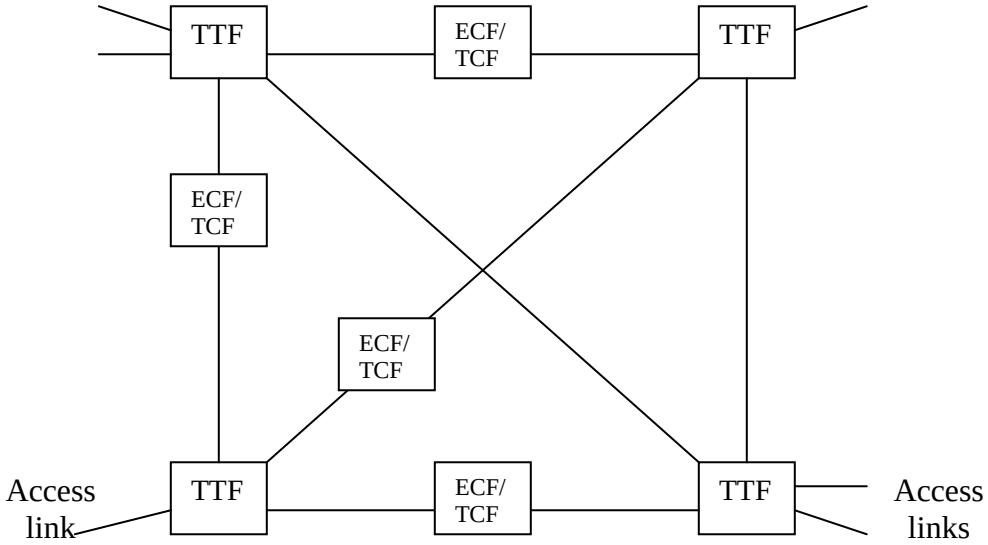

Figure 5: Split Horizon bridging with full mesh connectivity

Split-Horizon bridging is performed as follows: Each TTF maintains a bridging database controlling its bridging function. Each frame received by one of the TTF entities from an access-link is forwarded according to the bridging database of that TTF, to one, all, or some of the other TTF entities. Each copy is transmitted to one of the remote TTF entities through the direct trail leading to that remote TTF. Frames received by a TTF from one of the trails of the full-mesh, are forwarded by the TTF only to access-links.

With split-horizon bridging, the protection techniques discussed above are sufficient for protection of the MP2MP service, as long as each of the trails connecting the TTF entities is protected. A Split-horizon bridging subnet can serve as a subset of a larger bridged network, by connecting it to other bridging components, in this case its bridging elements may not be TTF entities, but ordinary ECF/TCF entities with split-horizon bridging capabilities.

The Spanning Tree Protocol is defined in [9], the Rapid STP is defined in [10]. These protocols provide protection in a network in which the TTF entities are connected in a partial mesh, and each of the TTF entities performs 802.1D compliant bridging between the links and trails connected to it (access-links as well as trails of its own layer). ECF/TCF entities through which the trails between the TTF entities pass may also perform 802.1D bridging. Observe that 802.1D bridging requires all links between the bridging entities to be bi-directional; therefore this scheme requires all trails between ECF/TCF entities that perform bridging to be bi-directional. 802.1D requires the bridging to be performed over a subset of the network that forms a spanning-tree of that network, and here is where STP and RSTP come to help, creating a spanning-tree of trails that participate in the bridged network, which spans all TTF and ECF/TCF entities implementing the service. STP requires fast aging or reset of the bridging databases in case of a change in the topology of the created spanning-tree.

As described in [1], STP-BPDUs may be:

- Processed at the UNI, in which case, the subscriber network becomes part of the network for which a single STP is calculated.
- Tunneled by the service, in which case, the service is perceived by the subscriber network as a single segment. In this case, subscriber STP can be created between its sites.
- Dropped at the UNI, in which case the subscriber should manually ensure that his network does not contain loops going through the service.

Note that tunneling and discarding also mean that an internal (MEN) STP can be created which is separated from the subscriber STP. In case tunneling is performed, the subscriber STP is then transparently tunneled through the MEN.

With the link-redundancy scheme, a single TTF attaches a bridged network of ECF or TCF entities (depending on the layer), using two point-to-point trails, which do not necessarily end at the same ECF/TCF on their other side. The TTF under discussion chooses at any time a single operational trail to work with. The TTF uses one of the mechanisms available in the technology of the specific layer (e.g. OA&M) for monitoring the operational status of the two trails. The TTF forwards frames received from its access-link to the chosen trail, and processes or forwards frames received from this trail to its access-links. Frames received from the other trail are dropped. When the TTF decides to change the trail, which is used for forwarding, it should inform the bridged network, to which it is attached that a topology-change happened, so the bridging entities in it can initiate fast aging or flush their database.

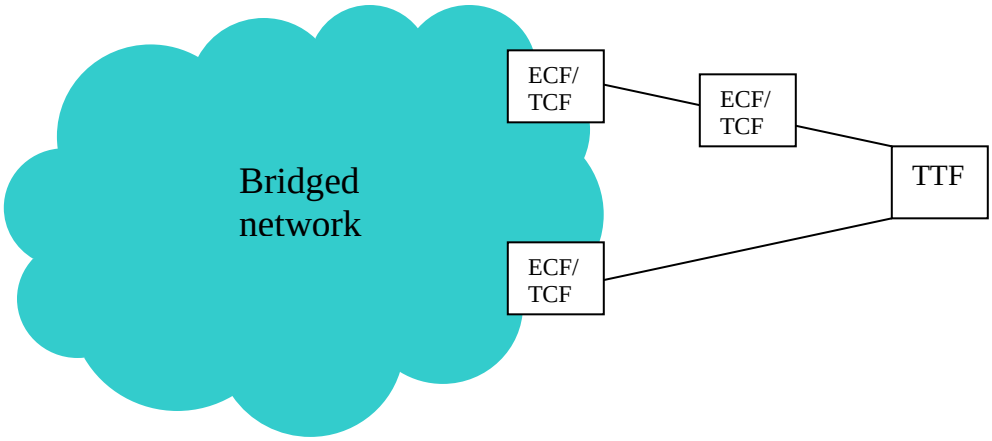

Figure 6: Link-redundancy

8.4 Link Protection based on Link Aggregation

For an Ethernet transport, Link Aggregation [11] allows one or more Ethernet links connecting the same two nodes to be aggregated into a Link Aggregation Group (LAG). A LAG logically behaves as a single link. The frames that each of the two nodes transmits through the LAG are distributed between the parallel links according to the decision of that node. The LAG distribution function should be such that it maintains the order of frames within each session using it. A LAG may be made of N parallel instances of full duplex point-to-point links operating at the same data rate. The Link Aggregation Control Protocol (LACP) is defined in [11], and is used by neighboring devices to agree on adding links to a Link Aggregation Group. The following figure is an example of a LAG topology:

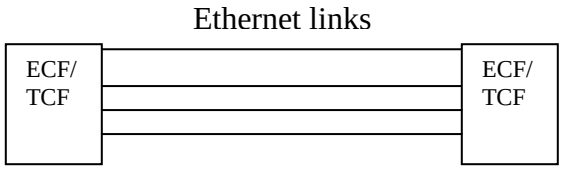

Figure 7: Link-aggregation

One of the features of a LAG setup is protection against failures of each of the links comprising the LAG. When one of these links fails, the nodes at the two sides of the LAG simply change their LAG distribution function to spread the traffic between the remaining links. When a link recovers, the LAG distribution function changes again to include it. An optional LA-marker protocol ensures that frame order is preserved during these changes.

The LAG scheme can be applied at the ETH-layer as well, by creating a number of point-to-point ETH-trails between two ECFs. The LACP can serve in this case as the OA&M procedure

detecting the failure and recovery of the ETH-trails, so that the LAG distribution function can be adapted accordingly.

8.5 Application Protection Constraint Policy (APCP)

Functionally APCP mediates between the subscriber and the MEN. It facilitates all the functions that are necessary to interpret service requests from MEN subscribers and to trigger services provisioning actions in the MEF transport network. Thus, a service request either through UNI or a management interface first comes to the APCP where the type of service and its parameters get interpreted. Then on behalf of the subscriber request, the APCP will in turn request (if validated) the MEN to provide the requested transport with desired capabilities/characteristics.

Details about the APCP support for different services as well as its interfaces facing the subscriber and the MEN sides will be defined in other MEF documents, such as the services document, the UNI specification, the protection scheme implementation agreements, etc.

9 Requirements for Ethernet Services protection mechanisms

The following are the requirements for the Metro Ethernet protection schemes. Every protection mechanism in the MEF protection framework is to be evaluated based on these requirements. The requirements are generic in the sense that they apply to all services supported by Metro Ethernet Networks (E-Line, E-LAN, CES).

9.1 Service-Related Requirements

9.1.1 Service Configuration Requirements

R-1. It **MUST** be possible for a subscriber to request different protection parameters for Ethernet services. The requested parameters **SHOULD** include the connectivity restoration time and SLS restoration time.

R-2. An EVC of an Ethernet service with SLS that requires protection **MUST** be protected along all ETH-trails from which it is comprised.

R-3. Protection parameters **MUST** be defined on the level of per-service or a group of services. End-to-end service protection **MAY** be implemented utilizing multiple mechanisms along the flow path.

9.1.2 Restoration Time Categories

Various applications require different connectivity and/or SLS restoration time. The restoration time (connectivity and/or SLS) that is required for a specific service is dependant on the needs of the application that the user plans to execute over that service.

R-4. It **SHOULD** be possible to request a connectivity and/or SLS restoration time of the network for each service. The following **SHOULD** be the restoration-time categories that the services can choose from for identifying the required connectivity restoration time and SLS restoration time:

- Sub 50ms restoration time.
- Sub 200ms restoration time.
- Sub 2 seconds restoration time.
- Sub 5 seconds restoration time.

Typical examples may be for CES applications to require sub50ms SLS restoration time, SLS restoration time of sub200ms is sufficient for certain CES applications, some real-time or semi real-time application may require also sub200ms SLS restoration time, sub 2 seconds connectivity restoration time ensures that a LAG implementation over the service using LACP in fast mode does not reconfigure due to the failure, TCP-based applications usually settle for sub 5 seconds connectivity restoration time, and connectivity restoration time of sub 5 seconds ensures that STP and RSTP do not start reconfiguring the network. These examples do not serve to categorize applications according to the restoration time, but only serve as examples for providing motivation.

9.2 Network Related Requirements

9.2.1 Protected failures

R-5. The protection mechanisms **SHOULD** be able to protect from any one of the failure types, for example:

1) Fail Condition (Hard link failure, e.g. LOS, LOF, etc.).
2) Degrade Condition (Soft link Failure, e.g. BER, CRC errors greater than a threshold).
3) Node failure

R-6. The protection mechanisms **MAY** protect from misconfigurations. Protection switching **MUST** not cause misconnections as a side effect of its own operation.

9.2.2 Degrade Condition Threshold

Degrade Condition is a status of a resource in which traffic transfer might be continuing, but certain measured errors (e.g., Packet loss, Bit Error Rate, etc.) have reached a pre-determined threshold.

R-7. With a protection scheme in which the SLS is preserved during failures, the predetermined threshold **MUST** be less than the maximum amount of packet-loss allowed according to the SLS of the services flowing through the resource. The provider **MAY** set less restrictive threshold on the resources if the SLS of all services flowing through the resource allows that.

One way for the network provider to meet the above requirement is to ensure that the packet loss commitment in SLS of the services is limited according to expected amount of packet loss allowed in the resources of the network.

It should be noted that depending on the protection scheme and layer other measurements than packet loss are used and a relation to packet loss might not be possible. After the protection

trigger (threshold crossing) it will take time (protection switching time) before the traffic is restored. During this time additional packet loss will occur.

9.2.3 Transport layer Protection Mechanisms Interaction

The MEN layering model allows layering within the TRANS layer. Therefore when one layer executes above another layer, both may belong to TRANS, or the upper one may be the ETH-layer and the lower-one may be a TRANS-layer. Protection in the APPS layer is beyond the scope of this document.

R-8. An upper layer protection mechanism **SHOULD** be designed to work in conjunction with lower layer transport protection mechanisms, such as SONET 1+1/1:1, RPR, etc, as available. Each protection mechanism that is allowed to execute in a network including these lower-layers **MUST** support configuration of the hold-off time such that the lower layer protection mechanism converges before the protection mechanism at the upper layer decides whether to take action. Note that the protection at the lowest layer doesn't need to support a hold-off timer.

In some cases, it **MAY** be required that the various layers will act independently. This depends on the protection policy for the network.

9.2.4 Protection Control Requirements

The following is a list of parameters and controls relevant for protection schemes.
1) Hold-Off Time.
2) Revertive/non revertive mode.
3) Reversion (Wait To Restore) Time.
4) Manual switch
5) Forced switch
6) Lockout
Items (1), (2), and (3) above are configuration of protection parameters, while (4) (5) and (6) are actually control mechanisms which control the current protection operation mode.

The requirement for support of a hold off timer is covered already in the above sub-section.

R-9. The protection at each (ETH/TRANS) layer **MAY** enable configuration of the Revertive/non revertive mode and Reversion (Wait To Restore) Time. These parameters and controls **MAY** be applied according to the protection policy. Protection **MAY** be applied in different layers of the MEN, and Revertive/non revertive mode and Reversion (Wait To Restore) Time **SHOULD** be applied to each mechanism separately.

R-10. The protection at each (ETH/TRANS) layer **SHOULD** enable configuration of at least one of the following controls:

- Manual switch
- Forced switch
- Lockout

Protection **MAY** be applied in different layers of the MEN, and the above three controls **SHOULD** be applied at each layer separately.

9.2.5 Bi-directional Switching

There are two variants of m:n protection type, one in which a protection resource can be used concurrently for protecting a number of working resources, in case a few of them fail at the same time. The other variant is where a protection resource is able to pass the traffic of a single working resource at a time.

R-11. In the case of a m:n protection scheme, in which the protection resource is able to pass the traffic of a single working resource at a time, and if the working and protection resources pass traffic in two directions, the bi-directional switching mechanism **MUST** be used for controlling the use of the protection resource.

The motivation is that bi-directional switching mechanism ensures that both directions of the same working resource switch concurrently to the protection resource. This ensures that the network does not get to a situation in which one direction of one working resource and another direction of another working resource are protected by the protection resource, so no single working resource is completely protected. An example for a mechanism that achieves this is the APS mechanism of SONET/SDH.

9.2.6 Robustness

R-12. Each protection mechanism **MUST** monitor protection standby resources for failures.

9.2.7 Backward Compatibility Requirements

R-13. When upgrading nodes in the network to support new protection mechanisms, these nodes **MUST** interoperate with nodes on the same network that are not yet upgraded to include these capabilities or schemes.

9.2.8 Network Topology

R-14. A Metro Ethernet Network may consist of different sub-network topologies. The protection framework **SHOULD** support these different topologies, although a specific scheme may be limited to few topologies.

R-15. The protection scheme **SHOULD** provide resource diversity such that the working and protection paths do not share a common resource in the network. The protection scheme **SHOULD** allow the required level of diversity to be according to the operator policy. The policy **MAY** require link diversity, node diversity, station diversity, fiber diversity, cable diversity, duct diversity, and geographical diversity. The operator **SHOULD** be able to control the required diversity either by controlling the network topology or the protection scheme and provisioning parameters.

R-16. To provide protection for a specific EVC, a protection scheme **MUST** protect each of the network resources within the network topology through which traffic of that EVC flows.

9.2.9 QOS

In many cases a tradeoff exists between efficient use of network resources and the extent of QoS preservation. The following requirement is helpful for operators for controlling this tradeoff.

R-17. Each protection method **SHOULD** enable the operator to define to what extent the original QoS is kept, up to full equivalent behavior.

9.2.10 Effect on user traffic

R-18. The protection mechanism **SHOULD** maintain the SLS requirements for loss of traffic and misordering during switching to protection and restoration events.

9.2.11 Management Requirements for Protection Schemes

R-19. A facility implementing a protection scheme **SHOULD** support a management interface that will expose to management applications the following parameters:

- All control parameters defined in 9.2.4
- Status of the working and protection paths
- Signaling of events that represent changes of status of the working and protection resources
- A failure event, which causes protection switch, **SHALL** cause an alarm to be sent by the involved Network Element(s) to the Element Management System.

10 Framework for Protection in the Metro Ethernet

10.1 Introduction

The key for the framework is the choice of protection mechanisms. Other aspects are important, but are handled elsewhere:

- The protection framework supports arbitrary transports. Some aspects of the transport layer, and its interaction with above layers, are discussed in Appendix A and Appendix B.
- We have asserted in Section 9 that protection solutions at each layer SHOULD be independent of the internal topology of the underlying layer.
- The APCP allows a subscriber to specify the parameters of the protection desired, this can be done through management-based configuration or the UNI.

Thus, our framework discussion focuses on the protection mechanisms and the failure detection mechanisms described above.

The model that has been developed has that property that it is open to new approaches and innovation within the layers. The description here reflects current discussion and can be expanded in the future.

10.2 MEF Protection Schemes

In this section, we outline some implementation approaches to providing MEF protection. The subsections currently developed regard OA&M-based End-to-end Path Protection, Aggregated

Line and Node Protection. MP2MP and LAG-based protection are already discussed in the former section. Other mechanisms can be added to this document at a later date. The details of the mechanisms will be found in ancillary Implementation Agreements.

Many protection mechanisms may be operating in the network at the same time. One reason is the existence of variety of services requiring different parameters of protection. Some of these protection mechanisms may offer protection at different layers of the network whereas some may offer different protection parameters such as restoration time, failure coverage, etc. at the same layer. However, operating many protection mechanisms at the same time in the network requires coherent interworking strategy so that trigger of multiple protection mechanisms for protecting the same traffic can be treated appropriately. Note that the use of multiple protection mechanisms for the same traffic may be desired because, for example, different protection mechanisms in the network may provide different failure coverage and restoration time. To ensure coherent operation, Section 9 requires that each protection mechanism that is allowed to execute in a network including these lower-layers MUST support configuration of the hold-off time such that the lower layer protection mechanism converges before the protection mechanism at the upper layer decides whether to take action.

10.2.1 OA&M-Based End-to-End Path Protection (EEPP)

This end-to-end path protection mechanism requires at least two paths to be created from the source node to the destination node for providing the same service. One of these paths is regarded as the primary path, and the others are redundant paths. The redundant paths are provisioned such that they are disjoint in nodes, links, and shared-risk links. In this way a failure of a single link or node will disconnect only one of the redundant paths at the most, so another can still be used and the service is maintained. This method can also be used to protect segments of the end-to-end path.

Possible modes of EEPP protection are: The 1:1 mode, in which two redundant paths are provisioned, but only one is used at a time; the n:1 mode, in which n+1 paths are provisioned, but only one is used at a time; the 1:n mode in which a single protection path is used for protecting n disjoined working paths; and the 1+1 protection mode, in which two paths are used concurrently - each data packet is sent along both paths and the sink node decides which to use.

An end-to-end OA&M protocol is used for sensing the availability of a path. With 1+1 protection, it is sufficient to have a one-way OA&M protocol. With 1:1, 1:n, or m:n protection-types, the source node is notified of the failure of the source to destination path. This means that the OA&M protocol **SHOULD** be a two-way protocol. An example for an OA&M protocol for MPLS can be found in ITU SG13 [4]. . Specific transports, like SONET/SDH, ATM, etc. have their own OA&M mechanisms that can be utilized for EEPP at the specific layer. Other OA&M mechanisms may also be used for the same purpose. Another variation of this scheme is when the end-to-end connectivity information is provided by out of band connections to a management station.

10.2.2 Aggregated Line and Node Protection (ALNP)

The ALNP scheme uses a set of local-protection tunnels for protecting each link and each node in the network. In this way an end-to-end SLS of fast (e.g., sub 50ms) protection is accomplished. The ALNP scheme can be applied to networks of any topology, provided that the topology provides a redundant-path for each of the network resources. ALNP is built to interoperate with any kind of link between the nodes implementing it. This can be a point-to-point physical link, a protected-transport (in which case protection-tunnels are required only for border nodes), a fast protected-subnet (again, only border nodes require ALNP protection), and virtual-links (in which case an OA&M procedure is required for indicating the state of the virtual-link and of its end-nodes).

10.2.3 Packet 1+1 End-to-End Path Protection

Packet 1+1 provides high reliability, hitless end-to-end path protection. Packets from and application flow receiving packet 1+1 service are dual-fed at the source side onto two disjoint paths. In the simplest case these paths can be node and link disjoint but in general may involve more complicated notion such as shared risk link groups. On the destination side one of the copies of the packet is selected from the two possible received copies. Note that the incoming packet is selected from any of the two disjoint paths irrespective of the path from which the last packet was selected. Thus packet 1+1 treats both paths as working. This is different from traditional transport 1+1 scheme where each path is designated as working or protection, and the packets are selected from the working until a detection of failure on the working causes a switching to the protection path. Thus, compared to traditional transport 1+1, packet 1+1 does not require explicit failure detection and protection switching. It also does not require any signaling. This allows the packet 1+1 service to recover from any failure that only affect one of the two disjoint paths providing the service instantaneously and transparently. Note these failures are protected irrespective of the layer, including physical, link TRANS and ETH layers, at which the failure occurred, as long as the failure is in the layer at which the mechanism is provided or in a layer below it. In other words, when the mechanism is provided at a specific layer, failures at that layer and in the layers below it are covered.

Preserving sequentiality requires defining a new protocol to include sequence numbers. Packet 1+1 requires that the source side assign the same but distinct identification to each dual fed packet. This can be easily achieved by e.g., assigning sequence numbers to packets. Each pair of duplicated packets will get the same sequence number but distinct from the other pairs of duplicate packets. Based on the carried sequence numbers the destination node is able to identify duplicate packets and select one of them. Note also that if a re-sequencing is not also provided, unerrored packets may be discarded to preserve sequentially.

10.2.4 Shared Mesh Protection

End-to-end shared protection scheme is targeted to provide guaranteed restoration while using minimal amount of protection bandwidth in a general mesh topology. Other end-to-end protection schemes either require a dedicated protection path for each primary path, such as 1+1 and 1:1, or provide a limited sharing of protection bandwidth, such as 1:N. Compared to them, shared protection scheme provides a very flexible sharing. Instead of dedicating protection

resources for every primary path in the network, a pool of protection resources can be set aside which can then be used for restoring the primary paths that get affected by the failure. Protection resources are allocated to allow restoration of all the protected traffic from any single possible failure in the network. For each protected primary path, the protection resources are allocated at the time of activation of the primary path. Arrival of a request to establish shared mesh restoration service between two nodes prompts computation of a pair of disjoint paths between them with two necessary constraints. First, sufficient bandwidth is allocated along the route of the primary path to accommodate the requesting traffic. Second, either already reserved protection bandwidth along the protection path is sufficient to guarantee restoration from any single failure along the primary route, or the available bandwidth along the protection path is allocated to be enough to accommodate the additional bandwidth needed for protecting the new primary path. Note that sharing is achieved by always first trying to accommodate a new request with already allocated protection resources. This can be achieved by keeping track, for each link in the network, the amount of resources that will be required to protect all the primary paths that will be switched to it after any single failure in the network. This information can be either maintained in a centralized fashion at a server or distributed to the nodes in the network. In the case where the information is distributed on the nodes, each node only needs to keep track of the amount of resources required on each of its incident links. Note that since it is well accepted and verified that the probability of multiple concurrent failures in most networks is small, the scheme has been described to protect from any single failure in the network. Protection from multiple failures can be achieved through a straightforward extension.

11 Requirements summary

The requirements listed in this specification are summarized below. The requirements were justified in the preceding sections.

1. It **MUST** be possible for a subscriber to request different protection parameters for Ethernet services. The requested parameters **SHOULD** include the connectivity restoration time and SLS restoration time. .

2. An EVC of an Ethernet service with SLS that requires protection **MUST** be protected along all ETH-trails from which it is comprised.

3. Protection parameters **MUST** be defined on the level of per-service or a group of services. End-to-end service protection **MAY** be implemented utilizing multiple mechanisms along the flow path.

4. It **SHOULD** be possible to request a connectivity and/or SLS restoration time of the network for each service. The following **SHOULD** be the restoration-time categories that the services can choose from for identifying the required connectivity restoration time and SLS restoration time:
    - Sub 50ms restoration time.
    - Sub 200ms restoration time.
    - Sub 2 seconds restoration time.

- Sub 5 seconds restoration time.

5. The protection mechanisms **SHOULD** be able to protect from any one of the failure types, for example:
   1) Fail Condition (Hard link failure, e.g. LOS, LOF, etc.).
   2) Degrade Condition (Soft link Failure, e.g. BER, CRC errors greater than a threshold).
   3) Node failure

6. The protection mechanisms **MAY** protect from misconfigurations. Protection switching **MUST** not cause misconnections as a side effect of its own operation.

7. With a protection scheme in which the SLS is preserved during failures, the predetermined threshold **MUST** be less than the maximum amount of packet-loss allowed according to the SLS of the services flowing through the resource. The provider **MAY** set less restrictive threshold on the resources if the SLS of all services flowing through the resource allows that.

8. An upper layer protection mechanism **SHOULD** be designed to work in conjunction with lower layer transport protection mechanisms, such as SONET 1+1/1:1, RPR, etc, as available. Each protection mechanism that is allowed to execute in a network including these lower-layers **MUST** support configuration of the hold-off time such that the lower layer protection mechanism converges before the protection mechanism at the upper layer decides whether to take action. Note that the protection at the lowest layer doesn't need to support a hold-off timer.

9. The protection at each (ETH/TRANS) layer **MAY** enable configuration of the Revertive/non revertive mode and Reversion (Wait To Restore) Time. These parameters and controls **MAY** be applied according to the protection policy. Protection **MAY** be applied in different layers of the MEN, and Revertive/non revertive mode and Reversion (Wait To Restore) Time **SHOULD** be applied to each mechanism separately.

10. The protection at each (ETH/TRANS) layer **SHOULD** enable configuration of at least one of the following controls:
    - Manual switch
    - Forced switch
    - Lockout

    Protection **MAY** be applied in different layers of the MEN, and the above three controls **SHOULD** be applied at each layer separately.

11. In the case in which the protection resource is able to pass the traffic of a single working resource at a time, and if the working and protection resources pass traffic in two directions, the bi-directional switching mechanism **MUST** be used for controlling the use of the protection resource.

12. Each protection mechanism **MUST** monitor protection standby resources for failures.

13. When upgrading nodes in the network to support new protection mechanisms, these nodes **MUST** interoperate with nodes on the same network that are not yet upgraded to include these capabilities or schemes.

14. A Metro Ethernet Network may consist of different sub-network topologies. The protection framework **SHOULD** support these different topologies, although a specific scheme may be limited to few topologies.

15. The protection scheme **SHOULD** provide resource diversity such that the working and protection paths do not share a common resource in the network. The protection scheme **SHOULD** allow the required level of diversity to be according to the operator policy. The policy **MAY** require link diversity, node diversity, station diversity, fiber diversity, cable diversity, duct diversity, and geographical diversity. The operator **SHOULD** be able to control the required diversity either by controlling the network topology or the protection scheme and provisioning parameters.

16. To provide protection for a specific EVC, a protection scheme **MUST** protect each of the network resources within the network topology through which traffic of that EVC flows.

17. Each protection method **SHOULD** enable the operator to define to which extent the original QoS is kept, up to full equivalent behavior.

18. The protection mechanism **SHOULD** maintain the SLS requirements for loss of traffic and misordering during switching to protection and restoration events.

19. A facility implementing a protection scheme **SHOULD** support a management interface that will expose to management applications the following parameters:
    - All control parameters defined in 6.2.4
    - Status of the working and protection paths
    - Signaling of events that represent changes of status of the working and protection resources
    - A failure event, which causes protection switch, **SHALL** cause an alarm to be sent by the involved Network Element(s) to the Element Management System.

12 Appendix A: Transport Protection

12.1 General

This appendix is an informative appendix that describes the motivation and examples of interaction between a transport layer and an upper layer, where each of the two layers includes protection mechanisms.

Modern Transport Networks use a layered paradigm. The different layers are independent of each other, and interconnect through well-defined Service Access Points (SAP). The SAP is defined using basic data and control exchanges. The advantage of this method is that each layer

may be administrated and maintained separately of the others, by different specialized entities within a Service Provider organization, or even by different organizations.

The implementer of each layer may select to use one out of many technologies, taking into consideration its specific needs as fulfilled by the candidate technology. As with other features, different technologies have different capabilities with respect to detection and restoration. To provide an overall protection strategy the capabilities of each layer is taken into consideration, and it is recommended that each layer allow the lower layers to try and fix the fault before it takes any protective action of its own. This is performed by using different restoration times (e.g. OA&M-based EEPP set to react slower than the protected-transport, in rare cases this is required for ALNP as well, in which case a hold-off timer is used).

The first part of this section will discuss interaction with a transport network capable of its own protection, the second part will list indications from currently available transport network that can be used when transport protection is not available (for example Ethernet transport) or is ignored (for example if the ALNP is configured to protect independently of lower protection availability).

12.2 Layered protection characteristics

Usually the characteristics of the protection related to the layer they are implemented are:

- Ø The lower the layer, the faster the protection
- Ø The higher the layer the longer the path that can be protected

As an example (Figures 8, 9 and 10), let us assume a network that transports ETH-layer over MPLS and the MPLS path transverses through a SONET/SDH or RPR ring, so we may have a layered Network of ETH/MPLS/RPR or ETH/MPLS/SONET. SONET/SDH & RPR methods will be able to protect very fast any failure occurring in the ring part of the Network (Figure 8), but they will be unable to provide any protection if the failure is in the MPLS path outside the ring (Figure 9). MPLS protection may be slower than SONET/SDH & RPR (e.g. OA&M-based EEPP) or of identical performance (e.g. ALNP), but it will be able to protect the whole MPLS path. In the same way the ETH layer will be able to protect the end-to-end ETH-service connectivity using a protection mechanism operating at the ETH-layer (Figure 10).

Following the desired feature of layer independency, each layer may include methods for detecting failures and restoring service, without the support of lower or higher layers. This does not preclude from a lower layer to inform a higher layer that it detected a failure.

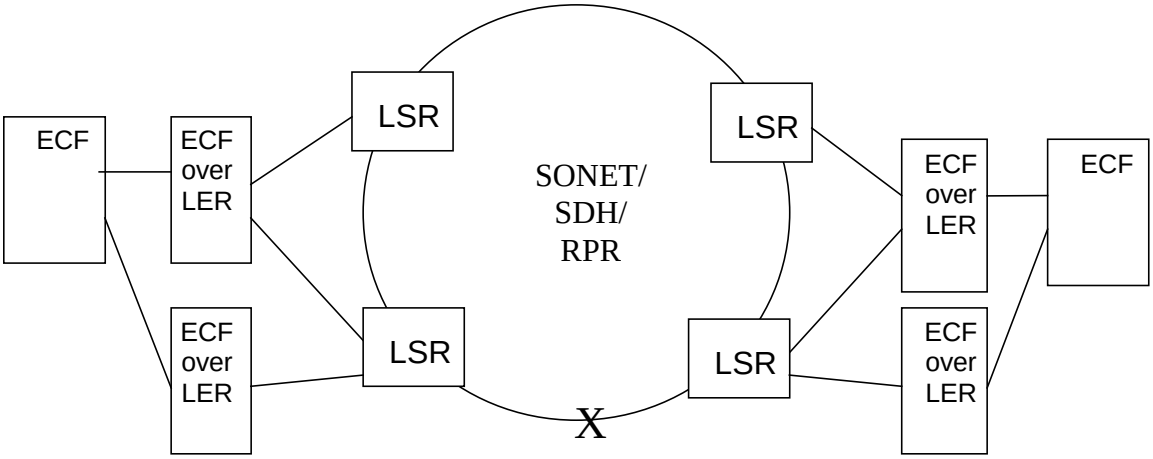

Figure 8: Failure that can be restored by SONET/SD or RPR

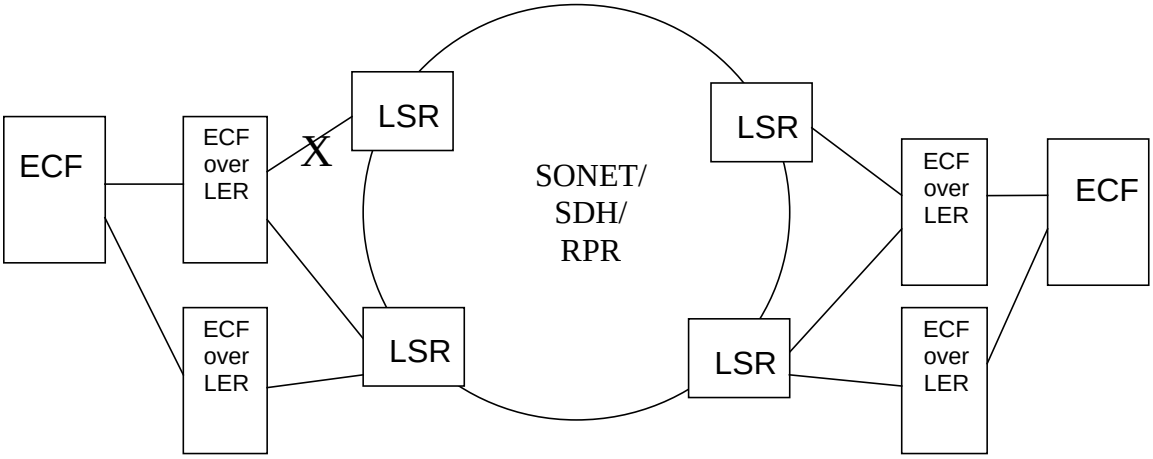

Figure 9: Failure cannot be repaired by SONET/SDH (BLSR, UPSR) or RPR; it can be repaired by MPLS

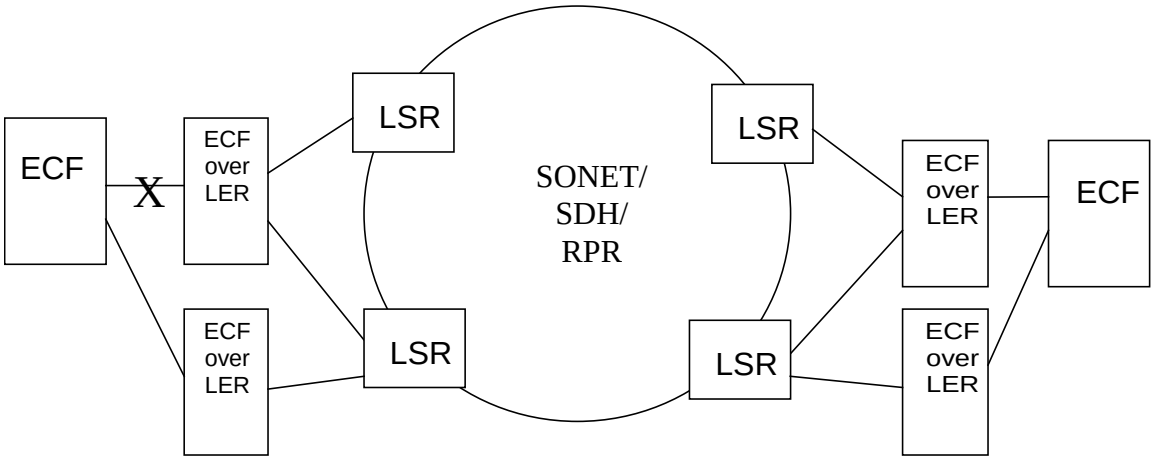

Figure 10: Failure can be restored by the ETH layer

12.3 Potential problems of protection interworking

Failing to synchronize and prioritize the operation of the protection at different layers may cause some undesirable network behavior. Following is an example of potential problems related to this situation.

Let us assume that no care is taken to synchronize and prioritize the protection mechanisms at different layers, and the simple network shown in Figure 11. Protection channels are shown in dotted lines, the connection between the Network Element (NE) executing ECF over SONET/SDH for the specific service under discussion and the Add Drop Multiplexer (ADM) is able to perform SONET Automatic Protection Switch (APS), and the NE is further able to use EEPP at the ETH-layer.

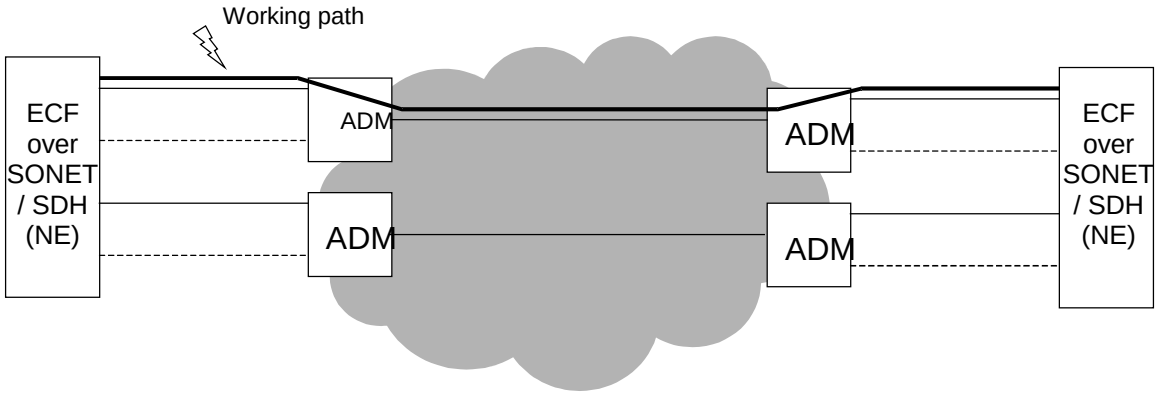

Figure 11: Simple network with protection at different layers

Let us assume that there is a failure as indicated, and that the EEPP is not synchronized with the APS, the result will be as shown in Figure 12. The EEPP protection path will be used, even though the APS was able to recover from the failure, and the working EEPP path (shown in dotted line) could be used.

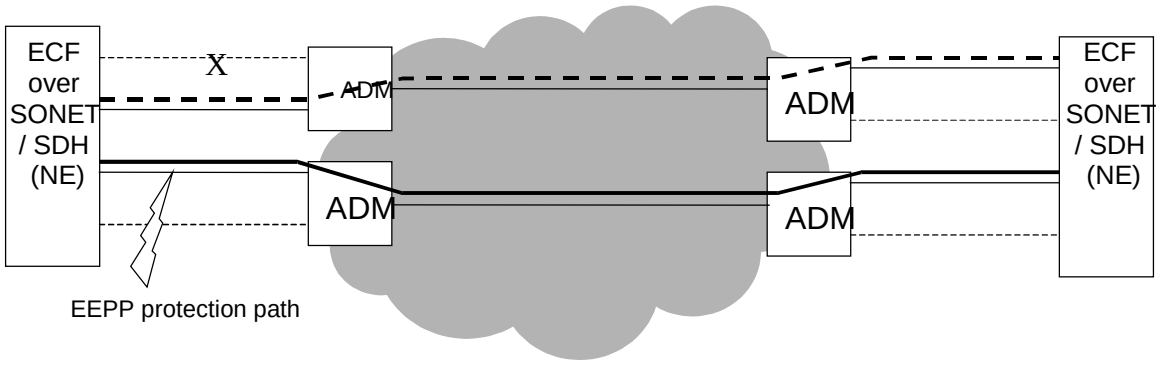

Figure 12: Protection by EEPP performed, but not required

If the EEPP procedure that is used is non-revertive, the network remains in this state and the resources in the EEPP protection path are overloaded unnecessarily. If the EEPP procedure is revertive, then a second switchover will take place, and the final result will be as shown in Figure 13.

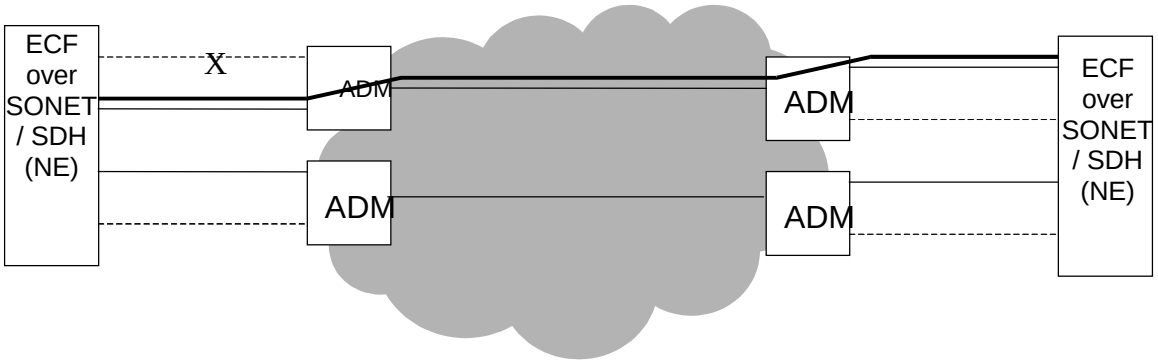

Figure 13: Final result for revertive switching

With the EEPP scheme, delaying the EEPP protection (e.g. using a hold-off timer), and allowing the APS mechanism to fix the problem will save the second switchover ("traffic hit") in the revertive mode, and better significantly improve resource utilization even for non-revertive mode. With ALNP, control exchange between the layers allows sub-50ms restoration. This example presents a very simplified case; more complex network topologies and layering could generate other issues. In cases in which the protected transport cannot give failure indication to the ALNP, a delay is required for interaction between ALNP and protected-transports as well.

12.4 Methods for internetworking between layers

If it is desired that a protection mechanism at a specific layer will allow the lower layers to attempt to fix the problem, and only if the lower layer is unable to complete the service restoration, the upper-layer will start its restoration procedure. When information about restoration status cannot be passed between the upper layer and the transport layer, allowing interoperability is a matter of delaying protection at the protection mechanism at the upper layer (e.g. using a hold-off timer).

13 Appendix B Transport Indications

As described in Section 9, if a transport services layer subnetwork does not have the means to do its own protection, or the protection capability is ignored as defined in the operator policy then the faults SHOULD be detected and repaired at a higher layer. The detection, however, may be based on instrumentation at the transport layer. The transport layer may provide enough statistics, alarms or fault notifications that the upper layer can use this to determine the status of the transport.

This appendix is an informative appendix that describes the available indications from various transport networks.

13.1 Optical transmission HW indications

Loss of light
Low light
More.

13.2 Ethernet HW indications

Loss of signal
Invalid signal
Auto-negotiation failure indications
Remote fault
Remote link loss

13.3 Ethernet-specific counters based decisions

CRC errors
Runts
Fragments
Alignment errors
Symbol errors
More.

13.4 SONET/SDH indications

SONET/SDH failure/degrade indications support both point-to-point (linear) and ring topologies and are delivered to client layers via well-known interfaces (via signaling or management planes), see [7] for more details.

13.5 RPR indications

RPR indications are ring wide based, and are sent to the upper layer after ring wise hierarchy of failure types and commands.

The native failure indications (before hierarchy) are:
1) Facility failures, like LOS, LOF etc.
2) Measurements based indications (considered soft or hard based on the appropriate threshold crossing):

- In SONET PHY BER measurement at the SONET layer.
- For Ethernet phys, the RPR MAC measure CRC errors on the received packets.

3) Equipment failure: RPR has built in node-to-node connectivity short timeout (order of milliseconds) as part of the fairness algorithm. In addition, RPR OAM continuity check (if configured) is able to detect node failure in order of seconds.

The hierarchy of failures is as follow (including management commands), from top to high:

- Forced Switch (FS) – Operator originated
- Signal Fail (SF) – Automatic

- LOS, LOF, L-AIS, BER (/CRC) above SF threshold in SONET/SDH (/Ethernet)
- RPR keep-alive timeout

· Signal Degrade (SD) – Automatic
  - BER (/CRC) above SD threshold in SONET/SDH (/Ethernet)
· Manual Switch (MS) – Operator originated
· Wait to Restore (WTR) – Automatic

In case RPR runs over SONET/SDH, part of these indications is supplied by SONET/SDH. In other cases, these indications are provided natively by RPR.

14 Appendix C (informative): Restoration Time Requirements derived from Customer Ethernet Control Protocols

Since the services provided by the MEN are Ethernet Services, they may need to transparently forward IEEE 802.1 and 802.3 control protocols, see [1]. The following are the control protocols, which periodically send control frames and therefore may be affected if the connectivity restoration time is too long. Protocols that are not related to 802.1 or 802.3 are not listed here. Also protocols that do not periodically send control packets (e.g. 802.1x) are not listed here, since such protocols are generally not affected by transient failures (as long as not informed of them).

[1] defines different modes of Ethernet services, in which the messages of different control protocols may be processed locally at the UNI, discarded at the UNI, or tunneled through the service. The discussion below refers only to the case in which the respective control frames are tunneled through the service. In the other two cases, the respective control packets are not forwarded by the service EVC, and are therefore not affected by its failures and their protection.

14.1 Spanning Tree Protocol and Rapid Spanning Tree Protocol

The Spanning Tree Protocol is defined in [9], the Rapid STP is defined in [10]. Both versions of the spanning tree protocol periodically send BPDUs once every hello-time, which is set to two seconds by default. Both use a timeout for aging of BPDU information that is calculated assuming that no more than three messages can be lost along the way from the root to the leafs of the created tree. For this reason, it is preferable to have a connectivity reaction time that is strictly less than eight seconds. The shorter the connectivity restoration time is, the less chances are that STP/RSTP will start rearranging the spanning-tree assuming that the link has failed.

14.2 Generic Attribute Registration Protocol

The Generic Attribute Registration Protocol (GARP) is defined in [9], and is used by layer2 devices to register and de-register attribute values with other layer2 devices.

GARP ensures that along each layer-2 segment, a LeaveAll BPDU is periodically sent once every LeaveAllTime (default is 10 seconds). When hearing this message, devices attached to that layer-2 segment need to re-register their attributes within LeaveTime (default is 600ms) or else their registration is canceled. The re-registering is performed by sending Join BPDUs at random periods of at most JoinTime (default is 200ms). This is performed twice for each attribute. If an

Ethernet service fails exactly after the LeaveAll message is sent, some of these re-registration messages may get lost, causing deregistration of the attribute until the next LeaveAllTime period.

In other words, GARP recovers after 10 seconds from periods of failure. If the connectivity restoration time is more than 400ms, interim side effects will occur. Below 400ms, the shorter the connectivity restoration time is, the lower the chances of interim side effects are.

14.3 Link Aggregation Control Protocol

The Link Aggregation Control Protocol (LACP) is defined in [11], and is used by neighboring devices to agree on adding links to a Link Aggregation Group, and to maintain packet ordering within each LAG.

The LACP protocol periodically monitors the links of the LAG, and can be configured to work in one of two modes, affecting the response time:

- Slow mode - in which an LACP message is sent once every 30 seconds, and timeouts after 90 seconds.
- Fast mode - in which an LACP message is sent once every 1 second, and timeouts after 3 seconds.

If the timeout expires for a specific link, the LACP reconfigures the LAG to use the other links only.

For an E-Line service that is used as part of a LAG that is controlled by LACP that works in fast mode, a connectivity restoration time of three seconds or more will result in the LACP taking action. For an E-Line service that is used as part of a LAG that is controlled by LACP that works in slow mode, a connectivity restoration time of 90 seconds or more will result in the LACP taking action.

15 References

[1] Metro Ethernet Forum, *Ethernet Service Model – Phase 1*, Technical Specification MEF 1.

[2] Bradner, S., *Key words for use in RFCs to Indicate Requirement Levels*, RFC 2119

[3] Sharma, V. and Hellstrand, F., *Framework for MPLS-based Recovery*, RFC 3469

[4] International Telecommunication Union, *OAM mechanism for MPLS networks*, Recommendation Y.1711

[5] International Telecommunication Union, *Protection Switching for MPLS Networks*, Recommendation Y.1720

[6] International Telecommunication Union, *Types and characteristics of SDH network protection architectures*, Recommendation G.841

[7] International Telecommunication Union, *Characteristics of synchronous digital hierarchy (SDH) equipment functional blocks*, Recommendation G.783

[8] International Telecommunication Union, *Network node interface for the Synchronous Digital Hierarchy (SDH)*, Recommendation G.707

[9] Institute of Electrical and Electronics Engineers, *Media Access Control (MAC) Bridges*, IEEE 802.1D-1998

[10] Institute of Electrical and Electronics Engineers, *Media Access Control (MAC) Bridges: Rapid Reconfiguration*, IEEE 802.1w-2001

[11] Institute of Electrical and Electronics Engineers, *Carrier sense multiple access with collision detection (CSMA/CD) access method and physical layer specifications*, IEEE 802.3-2002