TECHNICAL REPORT

TR-223

Requirements for MPLS over Aggregated Interfaces

(MPLSoAI)

Issue: 1
Issue Date: August 2012

Notice

The Broadband Forum is a non-profit corporation organized to create guidelines for broadband network system development and deployment. This Broadband Forum Technical Report has been approved by members of the Forum. This Broadband Forum Technical Report is not binding on the Broadband Forum, any of its members, or any developer or service provider. This Broadband Forum Technical Report is subject to change, but only with approval of members of the Forum. This Technical Report is copyrighted by the Broadband Forum, and all rights are reserved. Portions of this Technical Report may be copyrighted by Broadband Forum members.

This Broadband Forum Technical Report is provided AS IS, WITH ALL FAULTS. ANY PERSON HOLDING A COPYRIGHT IN THIS BROADBAND FORUM TECHNICAL REPORT, OR ANY PORTION THEREOF, DISCLAIMS TO THE FULLEST EXTENT PERMITTED BY LAW ANY REPRESENTATION OR WARRANTY, EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, ANY WARRANTY:

(A) OF ACCURACY, COMPLETENESS, MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, NON-INFRINGEMENT, OR TITLE;
(B) THAT THE CONTENTS OF THIS BROADBAND FORUM TECHNICAL REPORT ARE SUITABLE FOR ANY PURPOSE, EVEN IF THAT PURPOSE IS KNOWN TO THE COPYRIGHT HOLDER;
(C) THAT THE IMPLEMENTATION OF THE CONTENTS OF THE TECHNICAL REPORT WILL NOT INFRINGE ANY THIRD PARTY PATENTS, COPYRIGHTS, TRADEMARKS OR OTHER RIGHTS.

By using this Broadband Forum Technical Report, users acknowledge that implementation may require licenses to patents. The Broadband Forum encourages but does not require its members to identify such patents. For a list of declarations made by Broadband Forum member companies, please see http://www.broadband-forum.org. No assurance is given that licenses to patents necessary to implement this Technical Report will be available for license at all or on reasonable and non-discriminatory terms.

ANY PERSON HOLDING A COPYRIGHT IN THIS BROADBAND FORUM TECHNICAL REPORT, OR ANY PORTION THEREOF, DISCLAIMS TO THE FULLEST EXTENT PERMITTED BY LAW (A) ANY LIABILITY (INCLUDING DIRECT, INDIRECT, SPECIAL, OR CONSEQUENTIAL DAMAGES UNDER ANY LEGAL THEORY) ARISING FROM OR RELATED TO THE USE OF OR RELIANCE UPON THIS TECHNICAL REPORT; AND (B) ANY OBLIGATION TO UPDATE OR CORRECT THIS TECHNICAL REPORT.

Broadband Forum Technical Reports may be copied, downloaded, stored on a server or otherwise re-distributed in their entirety only, and may not be modified without the advance written permission of the Broadband Forum.

The text of this notice must be included in all copies of this Broadband Forum Technical Report.

Issue History

Issue Number	Approval Date	Publication Date	Issue Editor	Changes
1	21 August 2012	22 August 2012	Andrew Malis, Verizon Roman Krzanowski, Verizon Charles Rexrode, Verizon	Original

Comments or questions about this Broadband Forum Technical Report should be directed to info@broadband-forum.org.

Editors	Andrew Malis	Verizon
	Roman Krzanowski	Verizon
	Charles Rexrode	Verizon
IP/MPLS&Core WG Chairs	David Sinicrope	Ericsson
	Charles Rexrode	Verizon
Chief Editor	Michael Hanrahan	Huawei Technologies

TABLE OF CONTENTS

Executive Summary

TR-223, *MPLS over Aggregated Interfaces*, or MPLSoAI, defines a set of requirements for the use of MPLS over aggregated interfaces, such as IEEE 802.1AX *Link Aggregation* [1] or RFC 1990 *Multilink PPP (MLPPP)* [3].

It has become evident from lab testing and operational experience that while native MPLS should work fine over such aggregated interfaces by simply implementing the relevant specifications, such as 802.1AX and RFC 1990, the reality is that vendors have made some very different choices in their implementations. The intention of this document is to further specify the MPLSoAI interface so vendors can support multivendor interoperability and provide the required operational functionality.

1 Purpose and Scope

1.1 Purpose

Standards such as 802.1AX and MLPPP define the ability to aggregate or bundle together multiple link-layer interfaces to form a single logical interface from the viewpoint of higher layer protocols that are carried over the link layer. MPLS is one such higher layer protocol.

While MPLSoAI is supported in some shape or form by many vendors in their network equipment, the actual scope of the implementations differ. It has become evident from lab testing and operational experience that while native MPLS should work fine over such aggregated interfaces by simply implementing the relevant specifications, such as 802.1AX [1] and RFC1990 [3], the reality is that vendors have made some very different choices in their implementations.

Areas of major differences include the definitions of flows, allocation of flows among the members of the aggregated interface, and support for QoS, among others. These discrepancies are a major problem for operational and engineering teams that try to construct a network from different vendor equipment. The intention of this document is to further specify the MPLSoAI so interoperability can be supported and the required operational functionality be provided.

1.2 Scope

TR-223 defines a set of requirements for the use of MPLS over aggregated interfaces, such as IEEE 802.1AX *Ethernet link aggregation* and RFC 1990 *Multilink PPP (MLPPP)*.
TR-223 assumes that MPLS runs transparently over the bundled links resulting from the protocols mentioned above.

There is currently ongoing work in the IEEE on multi-chassis link aggregation and work in the IETF on using BFD over aggregated interfaces.
This work is currently outside the scope of TR-223.

2 References and Terminology

2.1 Conventions

In this Technical Report, several words are used to signify the requirements of the specification. These words are always capitalized. More information can be found be in RFC 2119 [4].

MUST	This word, or the term “REQUIRED”, means that the definition is an absolute requirement of the specification.
MUST NOT	This phrase means that the definition is an absolute prohibition of the specification.
SHOULD	This word, or the term “RECOMMENDED”, means that there could exist valid reasons in particular circumstances to ignore this item, but the full implications need to be understood and carefully weighed before choosing a different course.
SHOULD NOT	This phrase, or the phrase "NOT RECOMMENDED" means that there could exist valid reasons in particular circumstances when the particular behavior is acceptable or even useful, but the full implications need to be understood and the case carefully weighed before implementing any behavior described with this label.
MAY	This word, or the term “OPTIONAL”, means that this item is one of an allowed set of alternatives. An implementation that does not include this option MUST be prepared to inter-operate with another implementation that does include the option.

2.2 References

The following references are of relevance to this Technical Report. At the time of publication, the editions indicated were valid. All references are subject to revision; users of this Technical Report are therefore encouraged to investigate the possibility of applying the most recent edition of the references listed below.

Document		Title	Source	Year
[1]	802.1AX	*Link Aggregation*	IEEE	2008
[2]	ISO/IEC 10589	*Intermediate System to Intermediate System Intra-Domain Routing Exchange Protocol for use in Conjunction with the Protocol for Providing the Connectionless-mode Network Service (ISO 8473)*	ISO/IEC	2002
[3]	RFC 1990	*The PPP Multilink Protocol (MP)*	IETF	1996
[4]	RFC 2119	*Key words for use in RFCs to Indicate Requirement Levels*	IETF	1997
[5]	RFC 2328	*OSPF*	IETF	1998
[6]	RFC 2686	*The Multi-Class Extension to Multi-Link PPP*	IETF	1999
[7]	RFC 3032	*MPLS Label Stack Encoding*	IETF	2001
[8]	RFC 3209	*RSVP-TE: Extensions to RSVP for LSP Tunnels*	IETF	2001
[9]	RFC 3630	*Engineering (TE) Extensions to OSPF Version 2*	IETF	2003
[10]	RFC 3784	*Intermediate System to Intermediate System (IS-IS) Extensions for Traffic Engineering (TE)*	IETF	2004
[11]	RFC 4090	*Fast Reroute Extensions to RSVP-TE for LSP Tunnels*	IETF	2005
[12]	RFC 4124	*Protocol Extensions for Support of Diffserv-aware MPLS Traffic Engineering*	IETF	2005
[13]	RFC 4201	*Link Bundling in MPLS Traffic Engineering (TE)*	IETF	2005
[14]	RFC 4379	*Detecting Multi-Protocol Label Switched (MPLS) Data Plane Failures*	IETF	2006
[15]	RFC 4385	*Pseudowire Emulation Edge-to-Edge (PWE3) Control Word for Use over an MPLS PSN*	IETF	2006
[16]	RFC 4447	*Pseudowire Setup and Maintenance Using the Label Distribution Protocol (LDP)*	IETF	2006
[17]	RFC 4760	*Multiprotocol Extensions for BGP-4*	IETF	2006
[18]	RFC 5036	*LDP Specification*	IETF	2007

[19]	RFC 5085	*Virtual Circuit Connectivity Verification (VCCV): A Control Channel for Pseudowires*	IETF	2007
[20]	RFC 5881	*Bidirectional Forwarding Detection (BFD) for IPv4 and IPv6 (Single Hop)*	IETF	2010
[21]	RFC 5884	*Bidirectional Forwarding Detection (BFD) for MPLS Label Switched Paths (LSPs)*	IETF	2010
[22]	RFC 5885	*Bidirectional Forwarding Detection (BFD) for the Pseudowire Virtual Circuit Connectivity Verification (VCCV)*	IETF	2010
[23]	RFC 6391	*Flow-Aware Transport of Pseudowires over an MPLS Packet Switched Network*	IETF	2011

2.3 Definitions

The following terminology is used throughout this Technical Report.

ACH	Associated Channel Header
BFD	Bidirectional Forwarding Detection
BW	Bandwidth
CoS	Class Of Service
DA	Destination Address
DSCP	Differentiated Services Code Point
Flow	A set of frames transmitted from one end station to another, where all of the frames form an ordered sequence, and where the communicating end stations require the ordering to be maintained among the set of frames exchanged. [from 802.1AX section 3.8: conversation]
GAL	Generic Associated channel Label
IEEE	Institute of Electrical and Electronics Engineers
IETF	Internet Engineering Task Force
IP	Internet Protocol
ISIS	Intermediate System to Intermediate System
ISIS-DS-TE	ISIS – DiffServ Aware – Traffic Engineering
ITU-T	International Telecommunication Union – Telecom
LAG	Link Aggregation Group
LDP	Label Distribution Protocol
LSP	Label Switched Path
mBGP	Multiprotocol BGP

MEF	Metro Ethernet Forum
MIB	Management Information Base
MLPPP	Multilink PPP
MPLS	Multi Protocol Label Switching
MPLSoAI	MPLS over Aggregated Interfaces
MS-PW	Multi-segment Pseudowire
MTU	Maximum Transmission Unit
OAM	Operations, Administration and Management
OSPF	Open Shortest Path First
PCP	802.1Q Priority Code Point
PPP	Point to Point Protocol
PW	Pseudowire
QoS	Quality of Service
RFC	Request for Comments
RSVP-TE	Resource Reservation Protocol with Traffic Engineering Extensions
SA	Source Address
TC	Traffic Class
TE	Traffic Engineering
VCCV	Virtual Circuit Connectivity Verification
VLAN	Virtual LAN (IEEE 802.1Q)
VPLS	Virtual Private LAN Service
VPN	Virtual Private Network
VPWS	Virtual Private Wire Service

2.4 Abbreviations

This Technical Report uses the following abbreviations:

TR	Technical Report
WG	Working Group

3 Technical Report Impact

3.1 Energy Efficiency

Aggregated interfaces may be applied for energy efficiency. For example, during periods of low utilization, traffic may be aggregated onto a subset of the component links within a bundle, while other idle component links may be powered down or put into sleep mode to increase the energy efficiency of the overall bundle.

Specific details of using Aggregated Interfaces for energy efficiency are outside the scope of this document. Energy Efficiency is a work topic within the Broadband Forum.

3.2 IPv6

These requirements apply whether IPv4 or IPv6 is transported at the network layer above MPLS.

3.3 Security

TR-223 has no impact on security.

3.4 Privacy

TR-223 has no impact on privacy.

4 Reference Architecture

The scope includes two different types of interface multiplexing, IEEE 802.1AX [1] and MLPPP RFC 1990 [3]. The architecture and design principles of these interfaces are quite different. MPLSoAI can support either protocol.

4.1 IEEE 802.1AX Link Aggregation

IEEE 802.1AX Link Aggregation allows one or more links to be aggregated together to form a Link Aggregation. Link Aggregation does not support the following (as specified by Section 5.1.2/IEEE 802.1AX):

1. **Multipoint Aggregations**—The mechanisms specified in the 802.1AX clause referenced above does not support aggregations among more than two Systems.
2. **Dissimilar MACs**—Link Aggregation is supported only on links using the IEEE 802.3 MAC.
3. **Half-duplex operation**—Link Aggregation is supported only on point-to-point links with MACs operating in full duplex mode.

The position of Link Aggregation within the IEEE 802.3 architecture is specified in Section 5.1.3/IEEE 802.1AX. The link aggregation topology examples are provided in Figure A-1/IEEE 802.1AX.

The distribution algorithm selects the port used to transmit a given frame, such that the same port will be chosen for subsequent frames that form part of the same flow. This ensures the frame ordering.

4.2 RFC 1990 PPP Multilink Protocol

The PPP Multilink Protocol supports a method for splitting, recombining and sequencing datagrams across multiple logical data links. The bundled links can be comprised of different speed links.

Large packets are broken up into multiple segments sized appropriately for the multiple physical links. The PPP header, consisting of the Multilink Protocol Identifier and the Multilink header (with sequence number, etc.), is inserted before each section. (Thus the first fragment of a multilink packet in PPP will have two headers, one for the fragment, followed by the header for the packet itself).

MLPPP can work only for low speed links, partially because it uses sequence numbers to assemble segments.

MLPPP can support mobile backhaul traffic over multiple low speed interfaces (e.g., IP Base Station with E1/T1 interfaces and Cell Site Gateway with Ethernet to T1/E1 adaptation).

5 Requirements for IEEE 802.1AX

5.1 Link Aggregation protocol

Link Aggregation allows the establishment of a full duplex point-to-point link that has higher aggregate bandwidth than individual links that form the aggregation.

[R-1] Equipment implementing MPLSoAI to support Link Aggregation MUST support Link Aggregation as per IEEE 802.1AX [1].

[R-2] Equipment implementing MPLSoAI MUST support detaching a link from the aggregator as per Section 5.3.13/802.1AX.

[R-3] Equipment implementing MPLSoAI MUST support keepalive (periodic transmission) as per Section 5.4.13/802.1AX.

5.2 MPLS Support

[R-4] Equipment supporting MPLSoAI MUST support MPLS label stack encoding as defined by RFC 3032 [7].

[R-5] Equipment MUST support packets with at least five MPLS labels on an MPLSoAI interface.

[R-6] Equipment supporting MPLSoAI MUST support a common, configurable full Layer 2 MTU size across all links comprising the aggregated interface.

[R-7] MPLSoAI MUST support jumbo packets (up to at least 9K bytes).

[R-8] Packet order in the same flow MUST be maintained across the MPLSoAI by equipment when in a normal operational state, i.e., no failure.

[R-9] Equipment MUST support MPLS traffic engineering link bundling as per RFC 4201[13].

5.3 Load Balancing Support

[R-10] Equipment that implements MPLSoAI MUST support the ability to load balance across aggregated interfaces. Section A.2/802.1AX provides the port selection algorithm that preserves flow integrity.

[R-11] Equipment that implements MPLSoAI SHOULD include a configurable number of elements of the label stack (as specified in [R-5]) and/or adjacent IP header information, if present, in generating the hash.

[R-12] Equipment MUST NOT include Reserved labels in the label stack when generating **a** hash.

Note: It is recommended that equipment use a different hash-seed for MPLSoAI than it uses for ECMP to avoid polarization. This can also be achieved by the equipment supporting a different hash function for MPLSoAI than ECMP. The number of hash functions supported by equipment

is outside the scope of this specification.

Port Selection

Section A.2/802.1AX provides the port selection. One simple approach applies a hash function to the selected information to generate a port number. The keys chosen for the hash function depend on the packet type.

[R-13] An MPLSoAI interface MUST be supported across multiple interfaces in a single piece of equipment, regardless of physical implementation. For example, in a single card in a equipment, and across multiple cards in the same equipment.

[R-14] An MPLSoAI interface supports addition and deletion of links to the bundle.

1. An implementation MUST be capable of provisioning the addition and deletion of links to a bundle.
2. The addition or deletion of links not resulting from failure MUST have no loss to existing traffic.
3. Equipment SHOULD support Link Aggregation Control Protocol (LACP) as specified in Section 5.4/IEEE 802.1AX. Dynamic creation of aggregate bundles and dynamic association of a number of links to a bundle MAY be supported.

[R-15] An MPLSoAI interface MUST support the configuration of a minimum number of active links in the bundle before declaring failure of the entire bundle.

This requirement also applies to MLPPP.

[R-16] An MPLSoAI interface SHOULD support the configuration of the minimum available bandwidth in the bundle before declaring failure of the entire bundle.

5.4 QoS Support

[R-17] An MPLSoAI interface MUST be able to preserve the QoS designation (TC, DSCP, PCP) of the payload packet.

[R-18] An MPLSoAI interface MUST support setting the TC bits of the outer MPLS label independent of the QoS designation of the payload packet.

5.5 Control Protocol Support

If the equipment supports MPLS dynamic signaling, routing and Traffic Engineering, the following requirements apply:

[R-19] An MPLSoAI interface MUST support RSVP-TE (RFC 3209 [8]) and MPLS Fast Reroute (RFC 4090 [11]) for LSP signaling.

[R-20] An MPLSoAI interface MUST support LDP (RFC 5036 [18]) (including Targeted LDP) for LSP signaling.

[R-21] An MPLSoAI interface MUST support RFC 4447 [16] for Pseudowire signaling over both LDP and RSVP-TE signaled tunnels.

[R-22] An MPLSoAI interface MUST support mBGP (RFC 4760 [17]).

[R-23] An MPLSoAI interface MUST support ISIS [2], multi-instance ISIS, ISIS-TE (RFC 3784 [10]), and ISIS-DS-TE (RFC 4124 [12]).

[R-24] An MPLSoAI interface MUST support OSPF (RFC 2328 [5]), OSPF-TE (RFC 3630 [9]), and OSPF-DS-TE (RFC 4124 [12]).

5.6 OAM Support

[R-25] An MPLSoAI interface MUST support MPLS “ping” and MPLS “traceroute” functions as defined by RFC 4379 [14].

[R-26] An MPLSoAI interface MUST support VCCV functionality including Associated Channel Header (ACH) setup and signaling on an MPLSoAI interface, as per RFC 5085 [19].

[R-27] An MPLSoAI interface MUST support BFD over IP as defined in RFC 5881[20], BFD for MPLS as defined in RFC 5884 [21], and BFD over VCCV as defined in RFC 5085 [19].

NOTE: At the time of publication the IETF was working on several areas of OAM support.

5.7 Management

[R-28] An MPLSoAI interface SHOULD support the Management for Link Aggregation as per Section 6.3/802.1AX. Section 6.3.4/802.1AX Aggregation Port Debug information managed object class MAY be supported.

5.8 Security

[R-29] MPLSoAI MUST support MPLSoAI in compliance with Section 7.0/RFC 3032 [7] (Security Considerations).

6 Requirements for PPP Multilink Protocol RFC 1990

The PPP Multilink Protocol supports a method for splitting, recombining and sequencing datagrams across multiple logical data links. The bundled links can be comprised of different speed links.

[R-30] Equipment implementing MPLSoAI to support MLPPP MUST support RFC 1990 [3].

[R-31] Equipment implementing MPLSoAI and T1/E1 interfaces MUST support a bundle of multiple T1/E1 links.

[R-32] Equipment implementing MPLSoAI and T3/E3 interfaces MUST support a bundle of multiple T3/E3 links.

[R-33] Equipment implementing MPLSoAI and channelized STM-1/OC-3 interfaces MUST support a bundle of multiple VC-12 or Vt1.5 containers.

[R-34] Equipment implementing MPLSoAI SHOULD support the Multi-Class Extension to Multi-Link PPP as defined in RFC 2686 [6].

[R-35] Equipment implementing MPLSoAI MUST support the configuration of MLPPP packet fragmentation with sizes of 128 bytes, 256 bytes or 512 bytes.

[R-36] An MPLSoAI interface MUST support short sequence MLPPP fragment packet formats per section 3/RFC 1990 [3].

[R-37] An MPLSoAI interface SHOULD support long sequence MLPPP fragment packet formats per section 3/RFC 1990.

[R-38] An MPLSoAI interface SHOULD support in a MLPPP bundle with a number of links of different speeds.

[R-39] An MPLSoAI interface SHOULD support addition and deletion of link to MLPPP bundle with a minimum traffic loss.

[R-40] An MPLSoAI interface MUST support configuration of a common MTU value to be used across the MLPPP bundle.

6.1 QoS Support

QoS is supported per Section 5.4.

Appendix I. Load Balancing on Member Links of an Aggregated Interface

[INFORMATIVE]

IEEE 802.1AX is an Aggregated interface. Section A.2/802.1AX provides the port selection. One simple approach applies a hash function to the selected information to generate a port number.

A very important requirement when load balancing is that packets belonging to a given 'flow' must be mapped to the same port in an aggregated interface. This is to avoid the reordering of packets within the flow. What constitutes a flow varies with traffic and the packet criteria used to identify the flow, e.g., the Ethernet source and destination MAC addresses. In many cases, MPLS encapsulation may require fairly deep inspection of packets to find these criteria at transit LSRs.

PWs may be used to transport large volumes of IP traffic between Equipment. When the MPLS payload is a PW, an intermediate node has no information on the type of PW being carried in the packet. This limits the forwarder at the intermediate node to only being able to make a choice based on a hash of the MPLS label stack. In the case of a PW emulating a high bandwidth trunk, the granularity obtained by hashing the label stack may be insufficient for uniform load balancing over a LAG group.

Also, the mapping of flows to a particular component LAG link may not take into account the bandwidth of the flow being mapped or the current bandwidth usage of the members of the LAG.

IETF is working on solving the load balancing issue. The method for generating a flow label or entropy label should yield the maximum entropy given the source information available at the ingress of the PW or the LSP. Yielding maximum entropy helps achieve uniform distribution and address congestion within the LAG.RFC 6391 "*Flow-Aware Transport of Pseudowires over an MPLS Packet Switched Network*" defines better flow granularity and hashing for transport of pseudowires in a load balancing environment. The IETF is currently working on a similar entropy mechanism for LSPs and pseudowires.

Section 8/RFC 6391 [23] discusses the issues related to the LAG load distribution algorithms and the issue of congestion in a LAG component. The RFC proposes a solution to achieve uniform flow distribution over the LAG.

End of Broadband Forum Technical Report TR-223