MPLS Working Group
Internet Draft
Intended status: Standards Track
Expires: September 2012

Dave Allan, Ed.
Ericsson

March 2012

# Requirements and Framework for Unified MPLS Sub-Network Interconnection

draft-allan-mpls-unified-ic-req-frmwk-01

## Abstract

The definition of a transport profile for MPLS means that MPLS network architectures will emerge that combines both managed and control plane driven MPLS sub-networks and requires interconnection of same to achieve a unified MPLS architecture.

This document generalizes the problem of sub-network interconnect, discusses issues in general and suggests ways forward.

## Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC2119 [1].

## Status of this Memo

Copyright Notice

Table of Contents

## 1. Introduction

Networks that provide an end-to-end service infrastructure are typically deployed as multiple domains or sub-networks in order to scale. These different domains may be operated by different technologies using different control plane technology (e.g. management driven control plane (centralized) or distributed control plane).

Within the MPLS control plane there exist a number of functional behaviors that are typically associated with a single control protocol and function autonomously from the other control protocols. That is to say when a labeled layer and associated control protocol is overlaid on another, there is frequently no operational coupling between them. When two control protocols are operated side by side the same is true (e.g. ships in the night). An example of the former would be pseudo wires over a PSN. An example of the latter would be L2 and L3 virtual private networks (VPNs) sharing a common packet switched network (PSN).

The introduction of transport functions to MPLS and the intent to deploy, merge or otherwise combine networks that will concatenate sub-networks that have different operational characteristics and/or control planes (including management) introduces the requirement to integrate operations across multiple sub-networks. This frequently is manifested in requirements on nodes at sub-network boundaries and/or alignment of identifiers in order to construct a unified end-to-end MPLS based service plane.

By having a unified MPLS based service plane, a network operator is able to provide services across different MPLS domains instrumented with common management and control functionality in order to provide scalable service offerings on a common MPLS technology base.

This document is a framework that postulates models and functional requirements for sub-network interconnect to achieve a unified end to end MPLS based service plane or "Unified MPLS".

## 2. Conventions used in this document

### 2.1. Terminology

Binding: Is the mechanism for association and interconnection of label switched path (LSP) or pseudo wire (PW) segments that exist in different sub-networks.

Section: As per RFC 5960[17].

Segment: A part of a PW or LSP that has one or more contiguous sections in a common sub-network. This usage is as per RFC 6372[18].

Sub-network: A portion of a larger MPLS network that is operated by a single system and whose boundaries are set by the scope of the system. The system may be a control plane or management system. Similarly it could be a sub-layer stacked on another sub-network.

### 2.2. Acronyms

LIB - label information base

LSP - label switched path

MEG - Maintenance Entity Group

MEP - Maintenance Entity Group End Point

MIP - Maintenance Entity Group Intermediate Point

PSN - packet switched network

PW - pseudo wire

SPME - Sub-path Maintenance Entity

VPWS - Virtual Private Wire Service

VPMS - Virtual Private Multicast Service

## 3. Sub-Network Interconnect Scenarios

The combination of MPLS label swapping and label stacking makes it possible to consider a number of atomic interconnect scenarios, briefly summarized here:

Peer Model - is the scenario when an LSP or PW has segments in different sub-networks.

Segmentation Model - is the scenario when a client LSP or PW with common establishment and maintenance procedures has sections in separate sub-networks. This model can recurse arbitrarily.

Overlay Model - is the simplest case of the segmentation model in which a section of an LSP or PW is a distinct sub-network. This is currently the common deployment scenario for MPLS, and normally has minimal dependencies. In the specific case of MPLS & GMPLS, the requirements are examined in RFC 5146[21].

Termination Model - is the scenario where a non-MPLS or PW transfer function separates the sub-networks. The termination model logically isolates the sub-networks and is included simply for completeness.

## 4. Sub-Network Interconnect Mechanisms

There are two interconnect mechanisms considered. The first (Border node) postulates a node that spans two sub-networks and both likely operated by a single provider. The second (Border link) postulates a link connecting sub-networks of two distinct organizations within a provider or two distinct providers.

### 4.1. Border Node

A border node is one that implements, interconnects and interworks LSPs or PWs with segments or sections in different sub-networks. The interworking function at the border node will either terminate, swap or encapsulate labels.

### 4.2. Border Link

A border link is the case whereby two border nodes are connected back to back in an MPLS sub-network that consists of a single link. This can be a physical link or a logical link (e.g. an MPLS section).

## 5. Sub-network types

In the current MPLS architecture the following sub-network types exist:

### 5.1. Infrastructure sub-network types

The infrastructure sub-network types are:

MPLS-TD: Topology driven MPLS. LSP setup is via the LDP protocol [6] with path routing determined by the IGP. ECMP, merging and PHP are are included in the set of possible dataplane behaviors. LSPs are unidirectional only. P2mp and mp2mp LSPs are supported by mLDP [7].

MPLS-TE: LSP setup is via the RSVP-TE protocol[8] with path routing determined by a TE enabled IGP (e.g. OSPF-TE) according to bandwidth and QoS constraints. PHP is included in the set of dataplane behaviors. LSPs are unidirectional only. Bandwidth allocations, class of service or priority(PHB) are typically part of traffic handling.

Managed MPLS-TP: LSP setup is via management action. LSPs can be uni-directional, bi-directional or associated bi-directional. LSPs are purely connection oriented p2p or p2mp (where p2mp is unidirectional only).

Signalled MPLS-TP: LSP setup is via RSVP-TE GMPLS[9]. LSPs can be uni-directional, bi-directional or associated bi-directional. LSPs are purely connection oriented p2p or p2mp (where p2mp is unidirectional only).

And it is possible to consider

GMPLS for non-MPLS dataplanes:

### 5.2. Service Sub-network types

L3VPN: VPN setup is via the BGP protocol as per RFC 4364[10]. Note that this can be an IP-VPN (via IGP peering with the VRF) or an MPLS-VPN (via a combination of IGP and MPLS CP peering with the VRF).

L2VPN: VPN setup is via the PW Control Protocol per RFC 4447 (LDP in targeted mode) or BGP protocols. A broadcast domain is emulated which will have the effect of limiting sub-network interconnect to a single point to avoid loops.

VPWS: VPWS setup is via the PW Control Protocol per RFC 4447 (LDP in targeted mode).

VPMS: VPMS setup is currently an L2VPN work item .
[20].

## 6. Issues & Requirements

In theory, any ingress label can be mapped to one or more egress labels or label stack permutations via the ILM to NHLFE mapping defined in RFC 3031[2]. Further one carrier"s infrastructure layer may be a client of another carrier"s infrastructure. More considerations need to come into play in order to produce a tractable set of sub-network interworking scenarios. The following is a partial list of some of the issues to be considered and/or addressed:

### 6.1. Alignment of OAM functionality

Not all OAM encapsulations guarantee fate sharing with the LSP of interest across all of the sub-network types in MPLS. This not only means that failures may not be detected or detected in a timely manner, it also means that "false positives" are a possibility as failures may occur on the path taken by the OAM PDUs.

Any OAM encapsulation using a reserved label, e.g. the GAL[12], or Router Alert as used by VCCV type 2[13], or without a PW control word will not have predictable control over fate sharing with normal payload for any LSP or PW that has a section that transits a MPLS-TD sub network that implements ECMP[11]. Specifying that ECMP implementations exclude reserved labels from consideration would permit ECMP and LAG approaches that limit the sources of entropy to the label stack (e.g. FAT PWs [ref] or the entropy label [ref]) to be employed and be correctly and reliably instrumented by OAM that used a reserved label.

A separate issue is interconnecting sub networks where the LSPs have a different cardinality of end points (e.g. concatenating mp2p to p2p), implying a different number of maintenance entities than would be suggested by an implementation dimensioned to a single sub network"s characteristics.

### 6.2. OAM identifier mismatches

MEG, MEP, MIP and nodal addressing will not pose identifier mismatch problems. Where such problems will arise is in the use of RFC 4379 LSP Ping [3]. This is because LSP-PING uses identifiers associated with a specific sub-network type in the FEC stack as part of the processing to detect inconsistencies between the control plane and the data plane.

[Issue: The on demand CV draft provides for the Static LSP and Static PW TLVs for the FEC stack which allows intermediate nodes to validate the FEC stack against the MEG ID for the local MIP. The applicability for this should be generalized such that it can be used end to end across domains. This will also raise the problem of disseminating MEG information to non-transport sub networks as not all defined MPLS sub-network types use the current fields in the IP based LSP MEG ID]

### 6.3. OAM encapsulation

The MPLS architecture permits multiple OAM encapsulations that may or may nor have an IP header. Any interconnect mechanism needs to be able to align not only capability but encapsulations end to end. This

document assumes that translation of encapsulations by MIPs will not be specified or implemented.

## 6.4. Protection Mechanisms

An MPLS LSP or PW sub-network may be made resilient by any number of mechanisms. There are also three scenarios of note, end to end protection, end to end restoration and sub-network protection.

End to end protection offers minimal complications in sub-network interconnect as the interworking functions is common to that of the unprotected case, that is to say transit nodes do not participate in protection switching.

Sub-network protection is universally offered by the use of mechanisms that operate within the level such as detours [19] and may require label merging at the border node. Mechanisms that operate at nested MPLS label levels (e.g. SPMEs or FRR facility protection) have no impact on sub-network interconnect.

End to end restoration is a bit more complicated as it involves coordinating dynamic action between sub-networks.

It also becomes possible to consider sub-network restoration with many of the same considerations as path maintenance and re-sizing.

## 6.5. Label space management

The MPLS architecture has always been based around local administration of a node"s label space. As such mechanisms to partition the label space between multiple administrative entities is not currently supported and would be difficult to retrofit.

A consequence of this is that a border node is potentially required to provide labels from a common pool to both a control and management plane, e.g. a management system be required to obtain label values from the node prior to populating the LIB vs. being delegated a pool. This suggests that such a mechanism be carried forward for all managed nodes such that only a single mechanism need be implemented. However this is an implementation decision.

## 6.6. Path maintenance and re-sizing

It must be possible to make operational modifications to a path segment in a hitless fashion. The normal procedure for MPLS-TE is known as "make before break". This gives rise to two scenarios, the first is end to end "make before break", and the other is make before break confined to a sub-network with the border node as a pinned

waypoint. This means the design of the inter-sub-network binding information permit make before break modification of one segment of the LSP.

### 6.7. Sub-network migration

The practical considerations are documented in RFC 5950[4] and by reference RFC 5493[5].

### 6.8. (Non)Interworking of DP and CP notifications

Within the MPLS architecture there are techniques for propagating the status of adjacent sections of either a native service or PW section in both the data plane and the control plane. One example (documenting both) is RFC 6310[14].

When concatenating sub networks the interworking of dataplane fault notifications [15] or protection switching coordination [16] and control plane indications will not be possible. The reason is that data plane indications flow end to end on a labeled path therefore will not be visible to border nodes, a requirement to enable interworking of dataplane notifications with the control plane in any useful form.

When connecting a sub network restricted to data plane only notifications to a sub network that will support either dataplane or control plane notifications, the border node will be required to negotiate exclusive use of dataplane notifications in any control plane signaling during the path setup. This will have implications in both the interconnect data model, and potential enhancements to signaling.

## 7. Operational Decoupling

The objective of any sub-network interconnect solution is to decouple the operation of the interconnected systems in order to minimize any dependencies.

The sub-network interconnect must accommodate interconnecting LSPs and PWs with different establishment and persistency characteristics. This is determined by whether the LSP, PW or segment is provisioned or signaled, where from a persistency point of view, a provisioned entry is permanent and exists until removed by management action, while a signaled entity fate shares with a control plane adjacency and may come and go during the life time of the inter sub-network binding.

The state present at a border node to bind the LSP or PW spanning the sub networks together should exist independently of the characteristics of the LSPs or associated control or management planes.

This also requires a level of indirection such that the management action is decoupled from the mechanics of label assignment in each sub-network and may work with sub-network resiliency mechanisms.

So the state "connect whatever label from sub-network A associated with FOO to whatever label from sub-network B is associated with BAR" should be persistent.

## 8. Acknowledgments

Loa Andersson, Mach Chen, Eric Gray, David Sinicrope and Greg Mirsky contributed to the development of this document.

## 9. IANA Considerations

This document does not require IANA action.

## 10. Security Considerations

Sub-network interconnect in a single provider scenario does not introduce any new security exposures to the MPLS architecture that do not already exist.

Sub-network interconnect in a multi-provider scenario (e.g. border link) introduces a number of potential exposures, and requires a strong trust model for the co-ordination of the set up if interdomain LSPs. This is particularly true for the peer model. This is somewhat mitigated when operational decoupling techniques are employed as discussed in section 7, as the scope of what a provider can ask of a peer network is explicitly scoped.

## 11. References

### 11.1. Informative References

[1] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[2] Rosen et.al. Multiprotocol Label Switching Architecture, IETF RFC 3031, January 2001

[3] Kompella et.al. Detecting Multi-Protocol Label Switched (MPLS) Data Plane Failures, IETF RFC 4379, February 2006

[4] Mansfield et. al. Network Management Framework for MPLS-based Transport Networks, IETF RFC 5950, September 2010

[5] Caviglia, D., Bramanti, D., Li, D., and D. McDysan, "Requirements for the Conversion between Permanent Connections and Switched Connections in a Generalized Multiprotocol Label Switching (GMPLS) Network", RFC 5493, April 2009.

[6] Andersson et.al. "LDP Specification", IETF RFC 5036, October 2007

[7] Minei, I. et.al. "Label Distribution Protocol Extensions for Point-to-Multipoint and Multipoint-to-Multipoint Label Switched Paths", IETF RFC 6388

[8] Awduche et.al. "RSVP-TE: Extensions to RSVP for LSP Tunnels", IETF RFC 3209, December 2001

[9] Berger et.al. "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Functional Description", IETF RFC 3471, January 2003

[10] Rosen et.al. "BGP/MPLS IP Virtual Private Networks (VPNs)", IETF RFC 4364, February 2006

[11] Swallow et.al. "Avoiding Equal Cost Multipath Treatment in MPLS Networks", IETF RFC 4928, June 2007

[12] Bocci et.al. "MPLS Generic Associated Channel", IETF RFC 5586, June 2009

[13] Nadeau et.al "Pseudowire Virtual Circuit Connectivity Verification (VCCV) A Control Channel for Pseudowires", IETF RFC 5085, December 2007

[14] Aissaoui et.al. "Pseudowire (PW) Operations, Administration, and Maintenance (OAM) Message Mapping", IETF RFC 6310, July 2011

[15] Swallow et.al. "MPLS Fault Management OAM", IETF RFC 6427, November 2011

[16] Bryant et.al. "MPLS-TP Linear Protection", IETF RFC 6378, October 2011

[17] Frost et.al. "MPLS Transport Profile Data Plane Architecture", IETF RFC 5960, August 2010

[18] Sprecher et.al. "MPLS Transport Profile (MPLS-TP) Survivability Framework", IETF RFC 6372, September 2011

[19] Pan et.al. "Fast Reroute Extensions to RSVP-TE for LSP Tunnels", IETF RFC 4090, May 2005

[20] Kamite et.al., "Framework and Requirements for Virtual Private Multicast Service (VPMS)", IETF work in progress, draft-ietf-l2vpn-vpms-frmwk-requirements-04.txt, July 2011

[21] Kumaki et.al., Interworking Requirements to Support Operation of MPLS-TE over GMPLS Networks, IETF RFC 5146, March 2008

## Authors' Addresses

Dave Allan
Ericsson
Email: david.i.allan@ericsson.com