

Network Working Group
Internet Draft
Category: Standards Track

Don Fedyk, David Allan, Nortel
Greg Sunderwood, Bell Canada
Himanshu Shah, Ciena
Nabil Bitar, Verizon
Attila Takacs, Diego Caviglia, Ericsson

October 2006

GMPLS control of Ethernet
draft-fedyk-gmpls-ethernet-pbt-01.txt

Status of this Memo

By submitting this Internet-Draft, each author represents that any applicable patent or other IPR claims of which he or she is aware have been or will be disclosed, and any of which he or she becomes aware will be disclosed, in accordance with [Section 6 of BCP 79](#).

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at <http://www.ietf.org/ietf/lid-abstracts.txt>.

The list of Internet-Draft Shadow Directories can be accessed at <http://www.ietf.org/shadow.html>.

This Internet-Draft will expire in March 2007.

Copyright Notice

Copyright (C) The Internet Society (2006).

Abstract

Carrier Grade Ethernet transport solutions require the capability of flexible service provisioning and fast protection switching. Currently, Ethernet is being extended in IEEE to meet the scalability needs of transport networks.

The IETF specified GMPLS to control transport networks. To enable integration of Ethernet based transport solutions the extension of GMPLS control plane for Ethernet is of value.

This memo describes how a GMPLS control plane may be applied to Provider Backbone Transport (PBT) and how GMPLS can be used to configure VLAN-aware Ethernet switches in order to establish Ethernet P2P and P2MP MAC switched paths and P2P/P2MP VID based trees. This document assumes any standard changes to IEEE data planes will be undertaken only in the IEEE.

Conventions used in this document

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](#)].

Table of Contents

1. Introduction.....	5
2. Terminology.....	5
3. GMPLS Control of PBT Path creation and maintenance.....	6
3.1 Using a GMPLS Control Plane for Ethernet.....	7
3.2 Control Plane Network.....	7
3.3 Signaling.....	8
3.4 Ethernet Label.....	10
3.5 Ethernet Service.....	11
3.6 GMPLS Link Discovery.....	11
3.7 GMPLS Routing.....	12
3.8 Path Computation.....	12
3.8.1 Combinations of GMPLS Features.....	12
3.9 Addresses, Interfaces, and Labels.....	13
4. Specific Procedures.....	14
4.1 PT to PT connections.....	14
4.1.1 P2P connections with shared forwarding.....	14
4.1.2 Dynamic P2P symmetry with shared forwarding.....	15
4.1.3 Planned P2P symmetry.....	15
4.1.4 Path Maintenance.....	16
4.2 P2MP VID/DMAC Connections.....	16
4.2.1 Setup procedures.....	16
4.2.2 Maintenance Procedures.....	16
4.3 P2P/P2MP VID Trees.....	16
4.3.1 Setup Procedures.....	17
4.3.2 Maintenance procedures.....	17
4.4 OAM MEP ID and MA ID synchronization.....	17
4.5 Protection Paths.....	18
5. Error conditions.....	18
5.1 Invalid VID value for configured VID/DMAC range.....	18
5.2 Invalid VID value for configured VID range.....	18
5.3 Invalid MAC Address.....	18
5.4 Invalid ERO for Upstream Label Object.....	18
5.5 Invalid ERO for Suggested Label Object.....	19
5.6 Switch is not IVL capable.....	19
5.7 Switch is not SVL capable.....	19
5.8 Invalid VID in upstream label object.....	19
6. Deployment Scenarios.....	19
7. Security Considerations.....	19

8.	IANA Considerations.....	19
9.	References.....	19
9.1	Normative References.....	19
9.2	Informative References.....	20
10.	Author's Address.....	21
11.	Intellectual Property Statement.....	22
12.	Disclaimer of Validity.....	22
13.	Copyright Statement.....	22
14.	Acknowledgments.....	22
A 1.	Aspects of configuring Ethernet Forwarding.....	24
A 2.	Overview of configuration of VID/DMAC tuples.....	27
A 3.	Overview of configuration of VID port membership.....	29
A 4.	OAM Aspects.....	29

Fedyk et al.
Internet Draft

Expires March 2007
[draft-fedyk-gmpls-ethernet-pbt-01.txt](#)

Page 3

A 5.	QOS Aspects.....	30
A 6.	Resiliency Aspects.....	30
A 6.1	E2E Path protection.....	30

1. Introduction

Ethernet switches are increasing in capability. As a consequence the role of Ethernet is rapidly expanding in networks that were the domain of other technologies such as SONET/SDH TDM and ATM. The question of how Ethernet will evolve and what capabilities it can offer in these areas is still under development.

Operators are considering the deployment of Ethernet based transport solutions. The IEEE is working on amendments of VLAN-aware bridges (802.1Q) to meet scalability and service provisioning needs of operators. The work on 802.1ad Provider Bridges (PB) has already been finalized while the specification of 802.ah Provider Backbone Bridges (PBB) is expected to be ready in 2007. Parallel to the improvements of bridging functionalities standardization of 802.1ag Connectivity Fault Management (CFM) is also ongoing. CFM will equip bridged networks with service fault management and performance monitoring capabilities. In ITU-T Y.1731 work is ongoing to specify extensive OAM capabilities for Ethernet based on CFM. Moreover, in G.8031 Ethernet protection switching is being defined based on CFM's continuity check protocol. ITU-T G.8031 relies on p2p Ethernet paths configuration for working and protection traffic. P2p Ethernet paths are constructed using a p2p VLAN configuration between the head-end and tail-end of a protection segment. Note this is only a non-exhaustive list summarizing major activities pursuing Carrier Grade Ethernet transport.

The 802.1ad Provider Bridges and 802.1ah Provider Backbone Bridges are the respective amendments of the 802.1Q standard. The newly introduced functionalities add a hierarchical tunneling capability to Ethernet networks based on VLANs.

For Ethernet transport service provisioning, IEEE provides managed objects that can be statically configured through Network Management Systems and/or dynamically controlled through an Ethernet control Plane.

Provider Backbone Transport (PBT) is simply the data plane of Ethernet (802.1Q, 802.1ah) without an form of Spanning tree control plane. This document applies to PBT and is applicable to 802.1 when used for a suitable Pseudo wire service as described in this document.

The main purpose of this document is to specify a control plane for PBT that uses techniques for Ethernet.

[2. Terminology](#)

In addition to well understood GMPLS terms, this memo uses terminology from IEEE 802.1 and introduces a few new terms:

B-MAC	Backbone MAC
-------	--------------

Fedyk et al.	Expires March 2007	Page 5
Internet Draft	draft-fedyk-gmpls-ethernet-pbt-01.txt	

B-VID	Backbone VLAN ID
B-VLAN	Backbone VLAN
COS	Class of Service
C-MAC	Customer MAC
C-VID	Customer VLAN ID
C-VLAN	Customer VLAN
DMAC	Destination MAC Address
IVL	Independent VLAN Learning
MAC	Media Access Control
MP2MP	Multipoint to multipoint
PBB	Provider Backbone Bridge
PBT	Provider Backbone Transport
P2P	Point to Point
P2MP	Point to Multipoint
QOS	Quality of Service
SMAC	Source MAC Address
S-VID	Service VLAN ID
SVL	Shared VLAN Learning
VID	VLAN ID
VLAN	Virtual LAN

3. GMPLS Control of PBT Path creation and maintenance

PBT is an Ethernet connection technology, being specified in the IEEE, that can be controlled by configuration of static filtering entities [see [Appendix A](#)]. PBT paths are created switch by switch by simple configuration of Ethernet logical ports and assignment of PBT labels. We term a PBT path a form of Ethernet LSP (Eth-LSP). PBT paths may be configured by command line interface on the switches or coordinated from a management system. This memo proposes GMPLS as a mechanism to automate PBT paths.

One motive for using GMPLS over simple provisioning is GMPLS provides a reduction in the number of commands and an improvement in the coordination of commands required to establish and maintain an Eth-LSP. It also provides the capability for automation by dynamic modification of parameters, on-net/off-net path computation and automatic reaction to network changes without manual intervention. GMPLS utilizes per connection configuration and signaling both which reduce the operational overhead of establishing a path.

PBT uses the Ethernet data plane in its native form. When configuring a PBT path with GMPLS, the DMAC and VID are carried in a generalized label and are assigned hop by hop and it is invariant within a domain. PBT Eth-LSPs are by nature uni-directional since the DMAC must be inherently different in the two directions. The VID may be the same or different in each direction as it is only used to used to identify the path co-jointly with the DMAC. To be consistent with GMPLS terminology, paths are created first as an explicit route object (ERO) and data plane labels are assigned from the available

label pool at the destination switches. Each PBT label is a domain wide unique label, the VID/DMAC, for each direction.

Several attributes may be associated with an Eth-LSP, including:

- bandwidth requirements of the path. This can be used, for example, to request a fixed bandwidth path, where the committed information rate and peak information rate.
- priority level;
- preemption characteristics;
- protection/resiliency requirements;
- routing policy, such as an explicit route;
- policing requirements

Note GMPLS currently does not support unsymmetrical attributes in each direction for a bidirectional LSP. GMPLS control of PBT should

allow these parameters to be specified independently.

In addition to the above policies based on either under-subscription or over-subscription can be supported.

[3.1](#) Using a GMPLS Control Plane for Ethernet

GMPLS [[RFC3495](#)] has been adapted to the control of optical switches for the purpose of managing optical paths. GMPLS signaling is well suited to setup paths with labels but it does require an IP control plane and IP connectivity.

In many Ethernet deployment situations, the addition of a complete GMPLS control plane may be excessive for the switch technology or the network application. For this reason we consider partial application of GMPLS either complete functionality applied to a subset of the switches and/or partial functionality applied to some or all switches. For discussion purposes, we decompose the problem of applying GMPLS into the functions of Signaling, Routing, Link discovery and Path management. We can use some functions of GMPLS alone or in partial combinations. In most cases using all functions of GMPLS is less of an operational overhead than any partial combinations. Also, using only some components of GMPLS can lead to more provisioned overhead for some objects than a purely static system (see "Combinations of GMPLS Features").

[3.2](#) Control Plane Network

In order to have a GMPLS control plane, an IP control plane consisting of an IGP with TE extension needs to be established. This IGP views each hop as a terminated IP adjacency and should not be interpreted as a distinct routed subnet for the purpose of carrying IP bearer traffic. In other words IP is the control plane but the forwarding plane is not IP.

This IP control plane can be provided as a separate independent network (out of band) or integrated with the Ethernet switches.

If the IP control plane is a separate network, it may be provided as a Layer 2 connected LAN where the Ethernet switches are connected via a bridged network or HUB. It may also be provided by an external network that provides IP connectivity but in this case, the control topology of the GMPLS/Ethernet switches may not be the same topology as the physical data plane network.

If the IP control plane is integrated with the switches it may be provided by a bridged VLAN that uses the Data bearing channels of the network between adjacent nodes. This ties the fate of the controlled paths and the IP control plane links, which is not unlike the situation with MPLS networks or even some GMPLS optical networks.

3.3 Signaling

GMPLS signaling is well suited to the set up of PBT on Ethernet switches. GMPLS signaling uses either numbered or unnumbered links where the link is either explicitly IP addressed or associated with a switch loopback address. If LMP [[RFC4204](#)] is used, the creation of these unnumbered interfaces can be automated. If LMP is not used there is an additional provisioning requirement to add GMPLS link identifiers. For large-scale implementations LMP would be beneficial. As mentioned earlier, the primary benefit of signaling is the control of a path from an endpoint. GMPLS can be used to create bi-directional or unidirectional paths, bi-directional paths being the preferred mode of operation for numerous reasons (OAM consistency etc.). In this document we only consider bidirectional paths that share the same route/resources both for P2P and P2MP services.

Signaling enables the ability to dynamically establish a path and to adjust the path in a coordinated fashion after the path has been established. Signaling also improves multi-vendor interoperability over simple management since the signaling is standard and handles a number of dynamic functions. This allows the network to adapt to changing conditions or failures with a single mechanism. Signaling can be used for pure static configured paths as well.

To use GMPLS RSVP-TE signaling a few modifications are required. A new label is defined that contains the VID/DMAC tuple. On the initiating and terminating nodes, a function administers the VIDs associated with the SMAC and DMAC respectively. PBT is designed to be bidirectional and symmetric just like ethernet. Therefore in PBT the packet SMAC is the same as the DMAC used for the associated reverse PBT path and the DMAC is the same as the SMAC for the reverse PBT path.

To initiate a bi-directional VID/DMAC P2P or P2MP path, the initiator of the PATH message uses procedures outlined in [GMPLS-SIGNALING] possibly augmented with [[MPLS-P2MP](#)], it:

- 1) Sets the LSP encoding type to Ethernet.
- 2) Sets the LSP switching type to MAC [IANA to define].
- 3) Sets the GPID to Unknown (1) or Ethernet Multiplexed [IANA to define].
- 4) Sets the UPSTREAM_LABEL to the VID/SMAC tuple where the VID is administered from the configured VID/DMAC range. Downstream switches must use the SUGGESTED LABEL or return a path Error condition indicating why the label could not be used. Alternatively, if the optional LABEL SET object is implemented, the upstream switches can use this to specify the required label.

At intermediate switches the UPSTREAM_LABEL object and value is passed unmodified. The VID/SMAC tuple is used to create a static forwarding entry in the Filtering Database of bridges at each hop for the upstream direction. The port derived from the ERO and the VID and DMAC included in the label constitute the static entry.

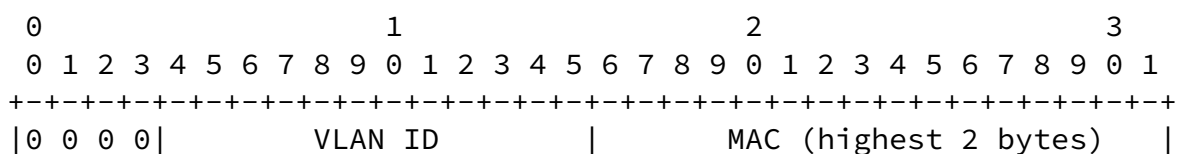
One capability of a connectionless Ethernet data plane is to reuse destination forwarding entries for packets from any source within a VLAN to a destination. When setting up point to point PBT connections for multiple sources sharing a common destination this capability can be preserved provided certain requirements are met. We refer to this capability as Shared Forwarding. Shared forwarding happens opportunistically when conditions are met as a local decision by label allocation at each end for the traffic to that end. To achieve shared forwarding, a Path computation engine should ensure the ERO is consistent with an existing path for the shared segments. If a path satisfies the consistency check, the upstream end of the signaling may choose to share an existing DMAC for the upstream traffic with an existing Eth-LSP. The consistency that the Eth-LSP share the same port and the paths of the Eth-LSP share one or more hops consecutively but once the paths diverge they must remain divergent. If no existing path has this behavior the path will be created unshared either by using another VID or another DMAC or both. In other words shared forwarding happens when paths share segments from the source and when the Upstream label is chosen to be the same as the existing path. Similarly for the downstream path shared forwarding happens when, an existing path that shares segments with the new paths ERO, viewed from the destination switch and when the downstream label is chosen to be the same and the existing path.

In this manner shared forwarding is a function that is controlled primarily by path calculation and in combination with the local label allocation at the end points of the path.

To initiate a P2MP VID path the initiator of the PATH message uses procedures outlined in [\[GMPLS-SIGNALING\]](#) and [\[MPLS-P2MP\]](#). A P2MP tree consists of a VID tree in the forward direction (from root to leaves) and a set of P2P paths running on identical paths from Tree to root in the reverse direction. VID labels with common MAC addresses are allocated in the forward direction and a single VID/DMAC label in the reverse direction:

- 1) Sets the LSP encoding type to Ethernet.
- 2) Sets the LSP switching type to L2SC.
- 3) Sets the GPID to unknown.
- 4) Set the technology specific information in the TSPEC to indicate domain-wide label.
- 5) Sets the UPSTREAM LABEL specified as a single VID/DMAC from the configured VID range.
- 6) VID translation may optionally be permitted on a local basis between two switches by a downstream switch replying with a VID/DMAC other than the SUGGESTED LABEL. The upstream switch then sets a VID translation on the port associated with the label to allow VID translation. This flexibility allows the tree to be constructed without having to worry about colliding with another tree using the same VID.

The Ethernet label is a new generalized label with a suggested format of:



```

+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|                                     MAC Address                                     |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+

```

The semantics of the new label type for a non-zero MAC address is that that the label is passed unchanged. This label is a domain wide label. This has similarity to the way in which a wavelength label is handled at an intermediate switch that cannot perform wavelength conversion, and is described in [[GMPLS-RSVP](#)].

These domain wide labels are allocated to switches that control the assignment of labels. This label space does not have to be globally unique because the labels are only valid within a single provider. When using configuration, a tool would have to perform a consistency check to make sure that label terminations were unique. When using GMPLS signaling it is assumed a unique pool of labels would be assigned to each switch. The DMAC addresses are domain wide unique and so is the combination of VID/DMAC. Should an error occur and a duplicate label be assigned to one or more switches GMPLS signaling procedures would allow the first assignment of the label and prevent duplicate label from colliding. If a collision occurs an alarm would be generated. In fact some of these procedures have been defined in GMPLS control of photonic networks where a lambda may exist as a form of domain wide label.

[3.5](#) Ethernet Service

Ethernet Switched Paths that are setup either by configuration or signaling can be used to provide an Ethernet service to customers of the Ethernet network. The Metro Ethernet Forum has defined some services in MEF.6 (e.g., Ethernet Private Line), and these are also aligned with ITU-T G.8011-x Recommendations. Of particular interest are the bandwidth profile parameters in MEF.10 and whose associated bandwidth profile algorithm are based on [[RFC4115](#)][RFC3270]. Consideration should be given to supporting these in any signaling extensions for Ethernet LSPs. This will be addressed in a future version of this specification.

[3.6](#) GMPLS Link Discovery

Link discovery is one of the building blocks of GMPLS. It is also a capability that has been specified for Ethernet in IEEE 802.1AB. Link discovery is optional but the benefits of running link discovery in large systems are significant. Link discovery reduces configuration and the possibility of errors in configuration as well as exposing misconnections. It is likely that a standard Ethernet

implementation would have 802.1AB functions. A recommendation is that standard 802.1AB could be run with an extension to feed information into an LMP [[RFC4204](#)] information model. LMP is a superset capability while 802.1AB has certain capabilities just for Ethernet. See Figure 3.

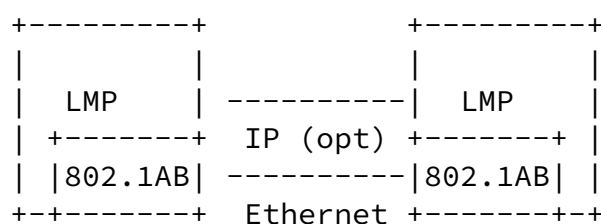


Figure 3 Link Discovery Hierarchy

[3.7](#) GMPLS Routing

GMPLS routing [[GMPLS-ROUTING](#)] is IP routing with the TLV extensions for the purpose of carrying around GMPLS TE information. The TE information is populated with TE resources from coordinated with LMP or from configuration if LMP is not available. The bandwidth resources of the links are tracked as Eth-LSPs are set up. GMPLS Routing is an optional piece but it is highly valuable in maintaining topology and distributing the TE database for path management and dynamic path computation.

[3.8](#) Path Computation

There has been a lot of recent activity in the area of path computation [[PATH-COMP](#)]. Originally MPLS path computation was performed locally in a TE database on a router. While this is non-optimal for situations where bandwidth is highly managed, it was acceptable when a few paths are required in a primarily connectionless environment; if a large number of coordinated paths are required, it is advantageous to have a more sophisticated path computation engine capable of optimizing the path routing of a sub network. The path computation could take the form of paths being computed either on a management station with local computation for rerouting or more sophisticated path computation servers.

[3.8.1](#) Combinations of GMPLS Features

The combinations of LMP, Routing, Signaling and Path computation can be all supported on a switch or a subset of GMPLS features can be supported.

Signaling is the minimal function that might be supported on a small switch. The ability to process Signaling messages with an ERO may be all that is desired on an end point. In this case the path computation would be performed off network.

Routing for GMPLS is the next important function since it provides the forwarding of signaling functions and transport of TE information. There is no requirement to provide full IP routing for data traffic, only hop by hop routing for the control plane. However it is possible to design switches without routing that could proxy their routing to other larger switches. In the proxied case, there would be little difference in the routing database but the small

switches would not have to perform routing operations. The information for the proxied routing might be configured or it might be data filled by an automated procedure. No new protocols are envisioned for the case where routing is proxied.

LMP is optional. The primary benefit of LMP in addition to 802.1AB is LMP's capability to optimize routing information for the purpose of link bundling on large switches. LMP has the capability to compress the information required to represent a large number of parallel resources automatically and reduce the amount of configuration.

[3.9](#) Addresses, Interfaces, and Labels

This specification uses an addressing scheme and a label space for the ingress/egress connection; the hierarchical GMPLS Switch Address/Port ID and the Ethernet VID/DMAC tuple or VID/Multicast MAC as a label space.

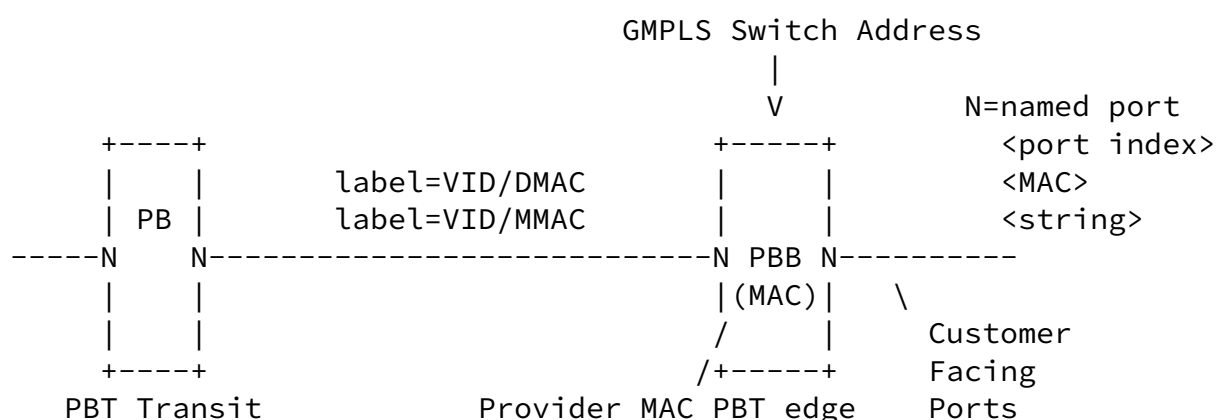


Figure 4 Ethernet/GMPLS Addressing & Label Space

Depending on how the service is defined and set up one or more of these identifiers may be used for actual setup. An Ethernet port name is common to both configured VID/DMAC, configured VID and bridging modes of operation. One implication of this is that a port index and a MAC address of a port on the switch may be effectively interchangeable for signaling purposes.

For a GMPLS based system, the GMPLS Switch Address/logical port is the logical signaling identifier for the control plane via which Ethernet layer label bindings are solicited. In order to create a point to point path an association must be made between the ingress and egress node. But the actual label distributed via signaling and instantiated in the switch forwarding tables identifies the upstream and downstream egress VID/DMAC of the PBT tunnel (see Figure 4). This label is typically an internal provider hidden domain wide label that is out of the locally administered label space.

GMPLS uses identifiers in the form of 32 bit numbers which are in the IP address notation but these are not IP addresses. An IP routing control plane for the propagation of TE information may be supported. The provider MAC port addresses are exchanged by the LLDP and by LMP if supported. However these identifiers are merely for link control and legacy Ethernet support and have local link scope. Actual label assignment is performed by the signaling initiator and terminator.

A particular port on a provider network switch would have:

- A MAC
- A 32 bit IPv4 Switch address or 128 bit IPv6 address plus 32 bit port Identifier
- One (or more) Mnemonic String Identifiers

This multiple naming convention leaves the issue of resolving the set given one of the port identifiers. On a particular node, mapping is relatively straightforward. The preferred solution to this is to use the GMPLS IP switch address for signaling resolution. In so doing, the problem of setting up a path is reduced to figuring out what switch supports an egress MAC address and then finding the corresponding GMPLS IP switch address and performing all signaling and routing with respect to the GMPLS switch address.

There are several options to achieve this:

- Provisioning

- Auto discovery protocols that carry MAC address
- Augmenting Routing TE with MAC Addresses
- Name Servers with identifier/address registration

This will be clarified in a subsequent version of this document.

4. Specific Procedures

4.1 PT to PT connections

The Data Plane for Ethernet has three key fields: VID, DMAC and SMAC. A connection instance is uniquely identified by the DMAC, the VID and the SMAC for the purpose of the provider network terminations. The VID and DMAC tuple identifies the forwarding multiplex at transit switches and a simple degenerate form of the multiplex is P2P (only one VID/DMAC/SMAC connection uses the VID/DMAC tuple).

This results in unique labels end to end. The data streams may merge, the forwarding entries may be shared but the headers are still unique allowing the connection to be de-multiplexed downstream.

4.1.1 P2P connections with shared forwarding

The VID/DMAC can be considered to be a shared forwarding identifier or label for a multiplex consisting of some number of P2P connections distinctly identified by the MAC VID/DMAC/SMAC tuple. The reason for using a shared forwarding entry is it reuses existing labels and

forwarding hardware. In some ways this is analogous to an LDP label merge but in the shared forwarding case the path control only the forwarding entry is reused.

VLAN tagged Ethernet packets include priority marking. Priority bits can be used to indicate class of Service (COS) and drop priority. Thus, traffic from multiple COSs could be multiplexed on the same ESP (i.e., similar to E-LSPs) and queuing and drop decisions are made based on the p-bits. This means that the queue selection can be done based on a per flow (i.e., ESP + priority) basis and is decoupled from the actual steering of the packet at any given node.

A switch terminating an ESP will frequently have more than one suitable candidate path and it may choose to share a forwarding entry.(common VID/DMAC , unique SMAC). It is a local decision of how this is performed but the best choice is a path that maximizes the shared forwarding.

The concept of bandwidth management still applies equally well with

shared forwarding. As an example consider a PBT edge switch that terminates an Ethernet LSP with the following attributes: bandwidth B1, DMAC D, SMAC S1, VID V. A request to establish an additional Ethernet LSP with attributes (bandwidth B2, DMAC D, SMAC S2, VID V) can be accepted provided there is sufficient link capacity remaining.

4.1.2 Dynamic P2P symmetry with shared forwarding

Similar to how a destination switch may select a VID/DMAC from the set of existing shared forwarding multiplexes rooted at the destination node, the originating switch may also do so for the reverse path. Once the initial ERO has been computed and the set of existing Ethernet LSPs that include the target DMAC have been pruned, the originating switch may select the optimal (by whatever criteria) existing shared forwarding multiplex for the new destination to merge with and offer its own VID/DMAC tuple for itself as a destination. This is identified via use of the UPSTREAM LABEL object.

4.1.3 Planned P2P symmetry

Normally the originating switch will not have knowledge of the set of shared forwarding paths rooted on the destination node.

Use of a Path Computation Server or other planning style of tool with more complete knowledge of the network configuration may wish to impose pre-selection of shared forwarding multiplexes to use for both directions. In this scenario the originating switch uses the SUGGESTED LABEL and UPSTREAM LABEL objects to indicate complete selection of the shared forwarding multiplexes at both ends. This may also result in the establishment of a new VID/DMAC path as the SUGGESTED LABEL object may legitimately refer to a path that does not yet exist.

4.1.4 Path Maintenance

Make before break procedures can be employed to modify the characteristics of a P2P Ethernet LSP. As described in [RFC3209], the LSP ID in the sender template is updated as the new path is signaled. The procedures (including those for shared forwarding) are identical to those employed in establishing a new LSP, with the extended tunnel ID in the signaling exchange ensuring that double booking of the associated resources does not occur.

Where individual paths in a protection group are modified, signaling procedures may be combined with Protection Switching (PS) coordination to administratively force PS switching operations such

that modifications are only ever performed on the protection path.

[4.2](#) P2MP VID/DMAC Connections

[4.2.1](#) Setup procedures

The group DMAC is administered from a central pool of multicast addresses and the VLAN selected from the configured VID/DMAC range. The P2MP tree is constructed via incremental addition of leaves to the tree in signaling exchange where the root is the originating switch (as per (MPLS-P2MP)). The multicast VID/DMAC are encoded in the suggested label object using the Ethernet label encoding.

Where a return path is required the unicast MAC corresponding to the originating interface and a VID selected from the configured VID/DMAC range is encoded as an Ethernet label in the upstream label object.

[4.2.2](#) Maintenance Procedures

Maintenance and modification to a P2MP tree can be achieved by a number of means. The preferred technique being to modify existing VLAN configuration vs. assignment of a new label and completely constructing a new tree.

Make before break on a live tree reusing existing label assignments requires a 1:1 or 1+1 construct. The protection switch state of the traffic is forced on the working tree and locked (PS not allowed) while the backup tree is modified. Explicit path tear of leaves to be modified is required to ensure no loops are left behind as artifacts of tree modification. Once modifications are complete, a forced switch to the backup tree occurs and the original tree may be similarly modified. This also suggests that 1+1 or 1:1 resilience can be achieved for P2MP trees for any single failure (switch on any failure and use restoration techniques to repair the failed tree).

[4.3](#) P2P/P2MP VID Trees

[4.3.1](#) Setup Procedures

The VID is administered from the central pool of VLAN IDs corresponding to the configured VID range. The P2MP VID tree is constructed via incremental addition of leaves to the tree in signaling exchange where the root is the originating switch as per [\[MPLS-P2MP\]](#).

Where (*,*) connectivity is to be configured a single VID is employed and encoded as an Ethernet label in the suggested label object with MAC address set to zero.

Where communication is to be constrained to root to leaves and leaves to root only, asymmetrical VID configuration is used with the suggested label object encoding the root to leaf VID and the upstream label object encoding the leaf to root VID.

[4.3.2](#) Maintenance procedures

Maintenance and modification to a P2P or P2MP VID tree can be achieved by a number of means. The preferred technique being to move traffic off the tree, modify the tree and then shift traffic back to the tree. This ensures that there are no transient loops in the tree that are artifacts of interactions of the GMPLS control plane, soft state and the Ethernet data plane.

Make before break on a live tree requires a 1:1 or 1+1 construct. The protection switch state of the traffic is forced on the working tree and locked (PS not allowed) while the backup tree is modified. Explicit path tear of leaves to be modified is required to ensure no loops are left behind as artifacts of tree modification. Once modifications are complete, a forced switch to the backup tree occurs and the original tree may be similarly modified. This also suggests that 1+1 or 1:1 resilience can be achieved for P2MP trees for any single failure (switch on any failure and use restoration techniques to repair the failed tree).

[4.4](#) OAM MEP ID and MA ID synchronization

The Maintenance end point IDs (MEP IDs) and maintenance association IDs for the switched path endpoints can be synchronized using the ETH-MCC (maintenance communication channel) transaction set once the switched path has been established.

MEPs are located at the endpoints of the Ethernet LSP. Typical configuration associated with a MEP is Maintenance Domain Name, Short Maintenance Association Name, and MA Level, MEP ID, and CCM transmission rate (when ETH-CC functionality is desired). As part of the synchronization, it is verified that the Maintenance Domain Name, Short Maintenance Association Name, MA Level, and CCM transmission rate are the same. It is also determined that MEP IDs are unique for each MEP.

Server MEPs can be considered at the intermediate points of the PBT network. Upon network failures (e.g. physical link failures), the Server MEPs can initiate the unicast AIS frames for each Ethernet LSP end-point that is present in the forwarding table. The only configuration required at the Server MEPs is the MA Level which should be the same as the MA Level configured at the Ethernet LSP MEPs.

Besides the unicast CCM and AIS functionality, the PBT MEPs can also offer the LBM/LBR and LMM/LMR functionalities for on-demand connectivity verification and loss measurement purposes.

[4.5](#) Protection Paths

When protection is used for path recovery it is required to associate the working and protection paths into a protection group. This is achieved as defined in [[RECOVERY_SIG](#)] using the ASSOCIATION and PROTECTION objects. Protection may be used for P2P VID/DMAC, P2MP VID/DMAC and P2P/P2MP VID configured modes of operation. The 'P' bit in the protection object indicates the role (working or protection) of the LSP currently being signaled.

If the initiating switch wishes to use G.8031 [[G-8031](#)] data plane protection switching coordination (vs. control plane notifications), it sets the N bit to 1 in the protection object. This must be consistently applied for all paths associated as a protection group.

If the terminating switch does not support G.8031, the error "Admission Control Failure/Unsupported Notification Type" is used.

[5.](#) Error conditions

The following errors have been identified as being unique to these procedures and in addition to those already defined. This will be addressed in a proper IANA considerations section in a future version of the document:

[5.1](#) Invalid VID value for configured VID/DMAC range

The originator of the error is not configured to use the VID value in conjunction with GMPLS signaling of VID/DMAC tuples. This may be any switch along the path.

[5.2](#) Invalid VID value for configured VID range

[5.3](#) Invalid MAC Address

The MAC address is out of a reserved range that cannot be used by then node which is processing the address.

[5.4](#) Invalid ERO for Upstream Label Object

The ERO offered has discontinuities with the identified VID/DMAC path in the UPSTREAM LABEL object.

[5.5](#) Invalid ERO for Suggested Label Object

The ERO offered has discontinuities with the identified VID/DMAC path in the SUGGESTED LABEL object.

[5.6](#) Switch is not IVL capable

This error may arise only in P2MP VID Tree allocation.

[5.7](#) Switch is not SVL capable

This error may arise only in P2MP VID Tree allocation.

[5.8](#) Invalid VID in upstream label object

The VID in the upstream label object for the "asymmetrical VID" P2MP tree did not correspond to the VID used in previous transactions.

[6.](#) Deployment Scenarios

This technique of GMPLS controlled Ethernet switching is applicable to all deployment scenarios considered by the design team [CCAMP-ETHERNET].

[7.](#) Security Considerations

The architecture assumes that the GMPLS controlled Ethernet subnet consists of trusted devices and that the UNI ports to the domain are untrusted. Care is required to ensure untrusted access to the trusted domain does not occur. Where GMPLS is applied to the control of VLAN only, the commonly known techniques for mitigation of Ethernet DOS attacks may be required on UNI ports.

[8.](#) IANA Considerations

New values are required for signaling and error codes as indicated. This section will be completed in a later version.

[9.](#) References

9.1 Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", [BCP 14](#), [RFC 2119](#), March 1997.

Fedyk et al. Expires March 2007 Page 19
Internet Draft [draft-fedyk-gmpls-ethernet-pbt-01.txt](#)

[CCAMP-ETHERNET] Papadimitriou, D. et.al, "A Framework for Generalized MPLS (GMPLS) Ethernet", internet draft, [draft-papadimitriou-ccamp-gmpls-ethernet-framework-00.txt](#) , June 2005

[GMPLS-SIGNALING] Berger, L. (editor), "Generalized MPLS -Signaling Functional Description", January 2003, [RFC3471](#).

[GMPLS-ROUTING] Kompella, K., Rekhter, Y., "Routing Extensions in Support of Generalized MPLS", [RFC 4202](#), October 2005

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

9.2 Informative References

[RFC4115] Aboul-Magd, O. et.al. "A Differentiated Service Two-Rate, Three-Color Marker with Efficient Handling of in-Profile Traffic", IETF [RFC 4115](#), July 2005

[G-8031] ITU-T Draft Recommendation G.8031, Ethernet Protection Switching.

[RFC3495] E. Mannie, Ed., "Generalized Multi-Protocol Label Switching (GMPLS) Architecture", [RFC 3495](#).

[IEEE 802.1ab] "IEEE Draft Standard for Local and Metropolitan Area Networks, Station and Media Access Control Connectivity Discovery".

[IEEE 802.1ag] "IEEE standard for Connectivity Fault Management", work in progress.

[IEEE 802.1ah] "IEEE standard for Provider Backbone Bridges", work in progress.

[RFC4204] Lang. J. Editor, "Link Management Protocol (LMP)" [RFC4204](#), October 2005

[MEF.6] The Metro Ethernet Forum MEF 6 (2004), "Ethernet Services Definitions - Phase I".

[MEF.10] The Metro Ethernet Forum MEF 10 (2004), "Ethernet Services Attributes Phase 1".

[RFC3270] Le Faucheur, F. et.al., "Multi-Protocol Label Switching (MPLS) Support of Differentiated Services" IETF [RFC 3270](#), May 2002.

[MPLS-P2MP] Aggarwal, R. Ed., "Extensions to RSVP-TE for Point to Multipoint TE LSPs", work in progress.

Fedyk et al. Expires March 2007 Page 20
Internet Draft [draft-fedyk-gmpls-ethernet-pbt-01.txt](#)

[MYERS] Myers et.al. "Rethinking the service model, scaling Ethernet to a million nodes", <http://100x100network.org/papers/myers-hotnets2004.pdf>.

[PATH-COMP] Farrel, A. et.al., "Path Computation Element (PCE) Architecture", work in progress.

[PWoPBT] Allan et.al., "Pseudo Wires over Provider Backbone Transport", [draft-allan-pw-o-pbt-01.txt](#), work in progress.

[RFC3985] Bryant, S., Pate, P. et al., "Pseudo Wire Emulation Edge-to Edge (PWE3) Architecture", IETF [RFC 3985](#), March 2005.

[RECOVERY_SIG] Lang et.al., "RSVP-TE Extensions in support of End-to-End Generalized Multi-Protocol Label Switching (GMPLS)-based Recovery ", work in progress.

[RFC3209] Awduche et.al., "RSVP-TE: Extensions to RSVP for LSP Tunnels, IETF [RFC 3209](#), December 2001.

[Y.1731] ITU-T Draft Recommendation Y.1731(ethoam), " OAM Functions and Mechanisms for Ethernet based Networks ", work in progress.

[10.](#) Author's Address

Don Fedyk
Nortel Networks
600 Technology Park Drive Phone: +1-978-288-3041
Billerica, MA, 01821 Email: dwfedyk@nortel.com

David Allan
Nortel Networks Phone: +1-613-763-6362
3500 Carling Ave. Email: dallan@nortel.com

Ottawa, Ontario, CANADA

Greg Sunderwood
Bell Canada
Suite 1500,
1066 West Hastings Street
Vancouver, BC, CANADA
V6E 2X1

Phone: +1-604-648-7770
Email: greg.sunderwood@gt.ca

Himanshu Shah
Ciena
35 Nagog Park,
Acton, MA 01720

Phone: 978-489-2196
Email: hshah@ciena.com

Nabil Bitar
Verizon,
40 Sylvan Rd.,
Waltham, MA 02451

Phone: (781) 466-2161
Email: nabil.n.bitar@verizon.com

Fedyk et al.
Internet Draft

Expires March 2007

Page 21

[draft-fedyk-gmpls-ethernet-pbt-01.txt](#)

Attila Takacs
Ericsson
1. Laborc u.
Budapest, HUNGARY 1037

Email: attila.takacs@ericsson.com

Diego Caviglia
Ericsson

Email: diego.caviglia@ericsson.com

11. Intellectual Property Statement

The IETF takes no position regarding the validity or scope of any Intellectual Property Rights or other rights that might be claimed to pertain to the implementation or use of the technology described in this document or the extent to which any license under such rights might or might not be available; nor does it represent that it has made any independent effort to identify any such rights. Information on the procedures with respect to rights in RFC documents can be found in [BCP 78](#) and [BCP 79](#).

Copies of IPR disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this specification can be obtained from the IETF on-line IPR repository at <http://www.ietf.org/ipr>.

The IETF invites any interested party to bring to its attention any

copyrights, patents or patent applications, or other proprietary rights that may cover technology that may be required to implement this standard. Please address the information to the IETF at ietf-ipr@ietf.org.

12. Disclaimer of Validity

This document and the information contained herein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

13. Copyright Statement

Copyright (C) The Internet Society (2006). This document is subject to the rights, licenses and restrictions contained in [BCP 78](#), and except as set forth therein, the authors retain all their rights.

14. Acknowledgments

Fedyk et al.
Internet Draft

Expires March 2007
[draft-fedyk-gmpls-ethernet-pbt-01.txt](#)

Page 22

The authors would like to thank Dinesh Mohan, Nigel Bragg, Stephen Shew and Sandra Ballarte for their extensive contributions to this document.

A 1. Aspects of configuring Ethernet Forwarding

Ethernet as specified today is a complete system consisting of a data plane and a number of control plane functions. Spanning tree, data plane flooding and MAC learning combine to populate forwarding tables and produce resilient any-to-any behavior in a bridged network.

Ethernet consists of a very simple and reliable data plane that has been optimized and mass produced. By simply disabling some Ethernet control plane functionality, it is possible to employ alternative control planes and obtain different forwarding behaviors.

Customer Bridge/	Provider Bridge	Provider Backbone Bridge
---------------------	--------------------	--------------------------------

```

C-MAC/C-VID-----802.1Q -----C-MAC-CVID
      S-VID-----802.1ad-----S-VID
            B-MAC---802.1ah---B-MAC
            B-VID---802.1ah---B-VID

```

Figure 1 802.1 MAC/VLAN Hierarchy

Recent works in IETF Pseudo Wire Emulation [[RFC3985](#)] and IEEE 802 are defining a separation of Ethernet functions permitting an increasing degree of provider control. The result is that the Ethernet service to the customer appears the same, yet the provider components and behaviors have become decoupled from the customer presentation and the provider has gained control of all VID/DMAC endpoints.

One example of this is the 802.1ah work in hierarchical bridging whereby customer Ethernet frames are fully encapsulated into a provider Ethernet frame, isolating the customer VID/DMAC space from the provider VID/DMAC space. Another example would be the direct transport of pseudo wires PWs ["Dry Martini" or PW over layer 2] where the Ethernet network fulfills the role of the PSN in the PWE architecture. In both cases the behavior of the provider's network is as per 802.1Q.

The Ethernet data plane provides protocol multiplexing via the ether type field which allows encapsulation of different protocols supporting various applications. More recently, the Carrier Ethernet effort has created provider and customer separation that enables another level of multiplexing. This in effect creates provider MAC endpoints in the Ethernet sub-network controlled by the provider. In this document we concentrate on the provider solutions and therefore subsequent references to VLAN, VID and MAC refer to those under provider control, be it in the backbone layer of 802.1ah or the PSN

layer of "Dry Martini". Also in the case where the Customer service is Ethernet, the Customer Ethernet service is the same native Ethernet service with functions such as bridging, learning and spanning trees all functioning over the provider infrastructure.

With the provider in exclusive control of their Ethernet sub-network and all customer specific state pushed to the edges of that sub-network, the ability of the provider to exploit latent Ethernet behavior is facilitated. One key capability sought is to overcome limitations, such as single spanning tree path for all traffic within a VLAN, imposed by bridging (see [[MYERS](#)] for a discussion).

Bridging offers a simple solution for any-to-any connectivity within a VLAN partition via the Spanning tree, flooding and MAC learning. Spanning tree provides some unnecessary capabilities for point to point services and since the Spanning tree must interconnect all MACs with the same VLAN IDs (VIDs) it consumes a scarce resource (VIDs). In this document we present that it is easier to modify Ethernet to scale engineered P2P services and this is the approach we take with PBT and PW over Ethernet. (The number of usable VLANs IDs in conventional Ethernet bridging is constrained to 4094, therefore the use of VLAN only configuration for all forwarding could be limited for some applications where large number of point to point connections are required.) This is because in Ethernet, each Spanning tree is associated with one or more VLAN IDs. Also Port membership in a VLAN is configured which controls the connectivity of all MAC interfaces participating in the VLAN.

The roots for PBT capability exist in the Ethernet management plane. The management of Ethernet switches provides for static configuration of Ethernet forwarding. The Ethernet Control plane allows for forwarding entries that are statically provisioned or configured. In this document we are expanding the meaning of "configured" from an Ethernet Control plane sense to mean either provisioned, or controlled by GMPLS. The connectivity aspects of Ethernet forwarding is based upon VLANs and MAC addresses. In other words the VLAN + DMAC are an Ethernet Label that can be looked up at each switch to determine the egress link (or links in the case of link aggregation).

In this document, we discuss, point to point (P2P) and point to multipoint (P2MP) connections via static configuration of VLAN/DMAC tuples. (MAC-only configuration is considered a degenerate case corresponding to VLAN zero).

This is a finer granularity than traditional VLAN networks since each P2P connection is independent. By provisioning MAC addresses independently of Spanning tree in a domain, both the VLAN and the VLAN/DMAC configured forwarding can be exploited. This greatly extends the scalability of what can be achieved in a pure Ethernet bridged sub network.

For compatibility and flexibility with existing Ethernet hardware, we preserve the global/domain wide uniqueness and semantics of MAC addresses as interface names or multicast group addresses. (In Ethernet overlap of MAC addresses across VLANs is allowed. However for PBT MAC addresses should be unique for all VLANs assigned to PBT. In many cases the MAC addresses can be out of the locally

administered space) We then redefine the semantics associated with administration and uses of VLAN values for the case of explicit forwarding such as you get with statically configured IVL (or SVL) Ethernet.

The result is a new architecture where configured VID + DMAC provide a forwarding table that is consistent with existing Ethernet switching. At the same time it provides domain wide labels that can be controlled by a common GMPLS control plane. This makes GMPLS control and resource management procedures ideal to create paths. The outcome is that the GMPLS control plane can be utilized to set up the following atomic modes of connectivity:

- 1) P2P connectivity and MP2P multiplexed connectivity based on configuration of unicast MAC addresses in conjunction with a VID from a set of pre-configured VIDs.
- 2) P2MP connectivity based on configuration of multicast MAC address in conjunction with a VID from a set of pre-configured VIDs. This corresponds to (Source, Group) or (S,G) multicast.
- 3) P2MP connectivity based on configuration of VID port membership. This corresponds to (S,*) or (*,*) multicast (where * represents the extent of the VLAN Tree).
- 4) MP2MP connectivity based on configuration of VID port membership (P2MP trees in which leaves are permitted to communicate). Although, we caution that this approach poses resilience issues (discussed in [section 5](#)) and hence is not recommended.

Items 1 and 2 above are restricted to "Independent VLAN Learning" capable Ethernet switches [802.1Q].

The modes above are not completely distinct. Some modes involve combinations of P2P connections in one direction and MP connectivity in the other direction. Also, more than one mode may be combined in a single GMPLS transaction. One example is the incremental addition of a leaf to a P2MP tree with a corresponding MP2P return path (analogous to a root initiated join).

In order to realize the above connectivity modes, a partition of the VLAN IDs from traditional Ethernet needs to be established. The partition allows for a pool of Ethernet labels for manual configuration and/or for GMPLS control plane usage. The VID partition actually consists of a "configured VID/DMAC range" and "configured VID range" since in some instances the label is a VID/DMAC and sometimes the label is a VID/Multicast DMAC.

A 2. Overview of configuration of VID/DMAC tuples

Existing Ethernet Switches may perform Independent VLAN Learning (IVL) based forwarding on the basis of a VID/DMAC tuple as described in 802.1Q. IVL is an example where the VLAN is partitioned and each is used as a unique filter for forwarding. In this document we build on that concept of IVL partitioning of the VID. The basic operation of an Ethernet switch is filtering on VID and forwarding on DMAC. The resulting operation is the same as performing a full 60 bit lookup (VID (12) + DMAC(48)) for point to point operations, only requiring uniqueness of the full 60 bits for forwarding to resolve correctly. We can call this an Ethernet domain wide label.

We have complete route freedom for each domain wide label (60 bit VLAN/DMAC tuple) and the ability to define multiple connectivity instances or paths per DMAC for each of the VIDs in the "configured VID/DMAC range".

We have preserved the semantics of MAC addresses, and simply broaden the potential interpretations of VLAN ID from spanning tree identifier to topology instance identifier. Therefore, we can concurrently operate both standard bridging and configured unicast/multicast operation side by side. We partition the VID space and allocate a range of VIDs (say 'n' VIDs) as only significant when combined with a configured DMAC address (the aforementioned "configured VID/DMAC range" of VIDs). We can then consider a VID in that range as an individual connectivity instance identifier for a configured P2P path terminating at the associated DMAC address. Or in the case of P2MP, a P2MP multicast tree corresponding to the destination multicast group address. Note that this is destination based forwarding consistent with how Ethernet works today. The only thing changed is the mechanism of populating the forwarding tables.

Ethernet MAC addresses are typically globally unique since the 48 bits consists of 24 bit Organizational Unique Identifier and a 24 bit serial number. There is also a bit set aside for Multicast and for local addresses out of the OUI field. We define domain wide as within a single organization, or more strictly within a single network within an organization. For provider MAC addresses that will only be used in a domain wide sense we can define MAC addresses out of either the local space or the global space since they both have the domain wide unique property. When used in the context of GMPLS, it is useful to think of a domain wide pool of labels where switches are assigned a set of MAC addresses. These labels are assigned traffic that terminates on the respective switches.

It is also worth noting that unique identification of source in the form of the SMAC is carried e2e in the MAC header. So although we have a 60 bit domain wide unique label, it may be shared by multiple sources and the full connection identifier for an individual P2P instance is 108 bits (SMAC, VID and DMAC). The SMAC is not

referenced in forwarding operations but it would allow additional context for tracing or other operations at the end of the path.

GMPLS signaling procedures can be designed to create the bi-directional path delegating label allocation of the combined VID/DMAC Label to the destination/source associated with the MACs for each direction of unicast forwarding. Creating P2P path is a well understood control plane requirement.

For multicast group addresses, the VID/DMAC concatenated label can be distributed by the source but label assignment (as it encodes global multicast group information) requires coordination within the GMPLS controlled domain.

As mentioned earlier, this technique results in a single unique and invariant identifier, in our case a VID/DMAC label associated with the path termination or the multicast group. There can be up to 4094 labels to any one MAC address. However, practically, from Ethernet network wide aspect, there would be only a handful of VLANs allocated for PBT. In addition, all 48 bits are not completely available for the MAC addresses. One way to maximize the space is to use the locally administered space. This is a large number for P2P applications and even larger when shared or multiplexed forwarding is leveraged. In practice, most network scaling requirements may be met via allocation of only a small portion of the VID space, to the configured VID/DMAC range. The result is minimal impact on the number of remaining bridging VLANs that can be concurrently supported.

In order to use this unique 60 bit label, we disable the normal mechanisms by which Ethernet populates the forwarding table for the allocated range of VIDs. When a path is setup, for a specific label across a contiguous sequence of Ethernet switches, a unidirectional connection is the functional building block for an Ethernet Label Switched path (Eth-LSP).

In P2P mode a bi-directional path is composed of two unidirectional paths that are created with a single RSVP-TE session. The technique does not require the VID to be common in both directions. However, keeping in line with regular Ethernet these paths are symmetrical such that a single bi-directional connection is composed of two unidirectional paths that have common routing (i.e. traverse the same switches and links) in the network and hence share the same fate.

In P2MP mode a bi-directional path is composed of a unidirectional tree and a number of P2P paths from the leaves of the tree to the root. Similarly these paths may have bandwidth and must have common

routing as in the P2P case.

There are a few modifications required to standard Ethernet to make this approach robust:

1. In Standard Ethernet, discontinuities in forwarding table configuration in the path of a connection will normally result in packets being flooded as "unknown". For configured operation (e.g.

Fedyk et al.
Internet Draft

Expires March 2007
[draft-fedyk-gmpls-ethernet-pbt-01.txt](#)

Page 28

PBT), unknown addresses are indicative of a fault or configuration error and the flooding of these is undesirable in meshed topologies. Therefore flooding of "unknown" unicast/multicast MAC addresses must be disabled for the "configured VID/DMAC range".

2. MAC learning is not required, and although it will not interfere with management/control population of the forwarding tables, since static entries are not overridden, it appears prudent to explicitly disable MAC learning for the configured VID/DMAC and VID range.

3. Spanning tree is disabled for the allocated VID/DMAC and VID range and port blocking must be disabled to achieve complete configured route freedom. As noted earlier, it is a control plane requirement to ensure configured paths are loop free.

All three modifications described above are within the scope of acceptable configuration options defined in IEEE802.1Q specification.

A 3. Overview of configuration of VID port membership

Procedures almost identical to that for configuration of P2P VID/DMAC tuples can also be used for the incremental configuration of P2MP VID trees. For the replication of forwarding in this case the label is common for the multipoint destinations. The MAC field is set to multicast address and is common to the multicast community. The VID is a distinguisher common to the multicast community. The signaling procedures are as per that for [[MPLS-P2MP](#)].

Since VID translation is relatively new and is not a ubiquitously deployed capability, we consider a VID to be a domain global value. Therefore, the VID value to be used by the originating switch may be assigned by management and nominally is required to be invariant across the network. The ability to indicate permissibility of translation will be addressed in a future version of the document.

A procedure known as "asymmetrical VID" may be employed to constrain connectivity (root to leaves, and leaves to root only) when switches

also support shared VLAN learning (or SVL). This would be consistent with the root as a point of failure.

A 4. OAM Aspects

Robustness is enhanced with the addition of data plane OAM to provide both fault and performance management.

For the configured VID/DMAC unicast mode of behavior, the hardware performs unicast packet forwarding of known MAC addresses exactly as Ethernet currently operates. The OAM currently defined, [802.1ag and Y.1731] can also be reused without modification of the protocols. However currently if the VID for PBT is different in each direction some modification of the OAM may be required.

Fedyk et al.
Internet Draft

Expires March 2007
[draft-fedyk-gmpls-ethernet-pbt-01.txt](#)

Page 29

An additional benefit of domain wide path identifiers for data plane forwarding, is the tight coupling of the 60 bit unique connection ID (VID/DMAC) and the associated OAM packets. It is a simple matter to determine a broken path or misdirected packet since the unique connection ID cannot be altered on the Eth-LSP. This is in fact one of the most powerful and unique aspects of the domain wide label for any type of rapid diagnosis of the data plane faults. It is also independent of the control plane so it works equally well for provisioned or GMPLS controlled paths.

Bi-directional transactions (e.g. ETH-LB) and reverse direction transactions (e.g. ETH-AIS) MAY have a different VID for each direction. Currently Y.1731 & 802.1ag makes no representations with respect to this.

For configured multicast VID/DMAC mode, the current versions of 802.1ag and Y.1731] make no representation as to how PDUs which are not using unicast addresses or which use OAM reserved multicast addresses are handled. Therefore this specification makes no representation as to whether such trees can be instrumented.

When configured VID mode of operation is used PBT can be forced to use the same VID in both directions, emulating the current Ethernet data plane and the OAM functions as defined in the current versions of 802.1ag and Y.1731 can be used with no restriction.

A 5. QOS Aspects

Ethernet VLAN tags include priority tagging in the form of the 802.1p priority bits. When combined with configuration of the paths via management or control plane, priority tagging produces the Ethernet equivalent of an MPLS-TE E-LSPs [RFC3270]. Priority tagged

Ethernet PDUs self-identify the required queuing discipline independent of the configured connectivity.

It should be noted that the consequence of this is that there is a common COS model across the different modes of configured operation specified in this document.

The actual QOS objects required for signaling will be in a future version of this memo.

A 6. Resiliency Aspects

A 6.1 E2E Path protection

One for One(1:1) protection is a primary LSP with a disjoint dedicated back up LSP. One plus one (1+1) protection is a primary LSP with a disjoint backup LSP that may share resources with other LSPs. One for One and One plus One Automatic Protection Switching strategies are supported. Such schemes offer:

Fedyk et al. Expires March 2007 Page 30
Internet Draft [draft-fedyk-gmpls-ethernet-pbt-01.txt](#)

- 1) Engineered disjoint protection paths that can protect both directions of traffic.
- 2) Fast switchover due to tunable OAM mechanisms.
- 3) Revertive path capability when primary paths are restored.
- 4) Option for redialing paths under failure.

Specific procedures for establishment of protection paths and associating paths into "protection groups" are TBD.

Note that E2E path protection is able to respond to failures with a number of configurable intervals. The loss of CCM OAM cells or ETH-AIS cells in the data plane can trigger paths to switch. In the case of CCM OAM cells, the detection time is typically 3.5 times the CCM interval plus the propagation delay from the fault.

