Network Working Group
Internet Draft
Intended status: Standards Track
Expires: September 8, 2011

D. Fedyk
Alcatel-Lucent
P.Ashwood-Smith
Huawei

March 8, 2011

# IS-IS Extensions Supporting IEEE 802.1aq Shortest Path Bridging
draft-ietf-isis-ieee-aq-05.txt

## Status of this Memo



## Copyright Notice

Abstract

802.1aq Shortest Path Bridging (SPB) is being standardized by the IEEE as the next step in the evolution of the various spanning tree and registration protocols. 802.1aq allows for true shortest path forwarding in a mesh Ethernet network context utilizing multiple equal cost paths. This permits it to support much larger layer 2 topologies, with faster convergence, and vastly improved use of the mesh topology. Combined with this is single point provisioning for logical connectivity membership, which includes point-to-point, point-to-multi-point and multi-point-to-multipoint variations. This memo documents the IS-IS changes required to support this IEEE protocol and provides some context and examples.

Table of Contents

1. Introduction................................................4
2. Terminology.................................................4
3. Conventions used in this document...........................5
4. 802.1aq Overview............................................6
   4.1. Multi Topology Support.................................8
   4.2. Data Path SPBM - Unicast...............................8
   4.3. Data Path SPBM - Multicast (Head End Replication).........9
   4.4. Data Path SPBM - Multicast (Tandem Replication)...........9
   4.5. Data Path SPBV Broadcast...............................11
   4.6. Data Path SPBV Unicast.................................12
   4.7. Data Path SPBV Multicast...............................12
5. SPBM Example................................................12
6. SPBV Example................................................15
7. SPB Supported Adjacency types...............................17
8. SPB IS-IS adjacency addressing..............................17
9. IS-IS Area Address and SYSID................................17
10. Level 1/2 Adjacency........................................17
11. Shortest Path Default Tie Breaking.........................17
12. Shortest Path ECT..........................................18
13. Hello (IIH) protocol extensions............................19
   13.1. SPB MCID sub-TLV......................................20
   13.2. SPB Digest sub-TLV....................................21
   13.3. SPB Base VLAN-Identifiers sub-TLV.....................24
14. Node information extensions................................25
   14.1. SPB Instance sub-TLV..................................26
      14.1.1. SPB Instance Opaque ECT-ALGORITHM sub-TLV..........29
15. Adjacency information extensions...........................30
   15.1. SPB Link Metric sub-TLV...............................30
      15.1.1. SPB Adjacency Opaque ECT-ALGORITHM sub-TLV.........31
16. Service information extensions.............................32
   16.1. SPBM Service Identifier and Unicast Address sub-TLV.....32
   16.2. SPBV Mac Address sub-TLV..............................33
17. Security Considerations....................................35
18. IANA Considerations........................................36
19. References.................................................37
   19.1. Normative References..................................37
   19.2. Informative References................................38
20. Acknowledgments............................................38
21. Author's Addresses.........................................38
22. Intellectual Property Statement............................39
23. Disclaimer of Liability....................................40

# 1. Introduction

802.1aq Shortest Path Bridging (SPB) [802.1aq] is being standardized by the IEEE as the next step in the evolution of the various spanning tree and registration protocols. 802.1aq allows for true shortest path forwarding in an Ethernet mesh network context utilizing multiple equal cost paths. This permits SPB to support much larger layer 2 topologies, with faster convergence, and vastly improved use of the mesh topology. Combined with this is single point provisioning for logical connectivity membership which includes point-to-point (E-LINE), point-to-multi-point (E-TREE) and multi-point-to-multipoint (E-LAN) variations.

The control protocol for 802.1aq is IS-IS [IS-IS] augmented with a small number of TLVs and sub TLVs. This supports two Ethernet encapsulating data paths, 802.1ad (Provider Bridges) [PB] and 802.1ah (Provider Backbone Bridges) [PBB]. This memo documents those TLVs while providing some overview.

Note that 802.1aq requires no state machine or other substantive changes to [IS-IS]. 802.1aq simply requires a new Network Layer Protocol Identifier (NLPID) and set of TLVs. In the event of confusion between this document and [IS-IS], [IS-IS] should be taken as authoritative.

# 2. Terminology

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

| | |
|---|---|
| 802.1ad | Provider Bridging (PB) - Q-in-Q encapsulation |
| 802.1ah | Provider Backbone Bridges (PBB), MAC-IN-MAC encapsulation |
| 802.1aq | Shortest Path Bridging (SPB) |
| Base-VID | VID used to identify a VLAN in management operations |
| B-DA | Backbone Destination Address 802.1ah PBB |
| B-MAC | Backbone MAC Address |
| B-SA | Backbone Source address in 802.1ah PBB header |
| B-VID | Backbone VLAN ID in 802.1ah PBB header |
| B-VLAN | Backbone Virtual LAN |
| BridgeID | 64 bit quantity = (Bridge Priority:16)<<48 \| SYSID:48 |
| BridgePriority | 16 bit relative priority of a node for tie breaking |
| C-MAC | Customer MAC. Inner MAC in 802.1ah PBB header |
| C-VID | Customer VLAN ID |
| C-VLAN | Customer Virtual LAN |
| DA | Destination Address |

```
   ECT-ALGORITHM  32 bit unique id of an SPF tie breaking set of rules.
   ECT-MASK       64 bit mask XORed with BridgeID during tie breaking.
   E-LAN          Bidirectional Logical Connectivity between >2 UNIs.
   E-LINE         Bidirectional Logical Connectivity between two UNIs.
   E-TREE         Asymmetric Logical Connectivity between UNIs.
   FDB            Filtering Database: {DA/VID}->{next hops}
   I-SID          Logical Grouping Identifier for E-LAN/LINE/TREE UNIs.
   LAN            Local Area Network
   LSDB           Link State Database
   LSP            Link State Packet
   MAC-IN-MAC     Ethernet in Ethernet framing as per 802.1ah[PBB]
   MDT            Multicast Distribution Tree
   MMRP           Multiple Mac Registration Protocol 802.1ak[MMRP]
   MT-ISIS        Multi Topology IS-IS as used in [MT]
   MT             Multi Topology. As used in [MT]
   MT-ID          Multi Topology Identifier (12 bits). As used in [MT]
   NLPID          Network Layer Protocol Identifier: IEEE 802.1aq= 0xC1
   Q-in-Q         Additional S-VLAN after a C-VLAN (802.1ad)[PB]
   PBB            Provider Backbone Bridge - forwards using PBB
   Ingress Check  Source Forwarding Check - drops misdirected frames
   (S,G)          Source & Group - identity of a source specific tree
   (*,G)          Any Source & Group - identity of a shared tree
   SA             Source Address.
   SPB            Shortest Path Bridging - generally all of 802.1aq.
   SPB            Shortest Path Bridge - device implementing 802.1aq.
   SPB-instance   Logical SPB instance correlated by MT-ID.
   SPBM           Device implementing SPB MAC mode
   SPBV           Device implementing SPB VID mode
   SPT            Shortest Path Tree computed by one ECT-ALORITHM
   SPT Region     A set of SPBs with identical VID usage on their NNIs
   SPSourceID     20 bit identifier of the source of multicast frames.
   SPVID          SPBV: a C-VLAN or S-VLAN that identifies the source.
   UNI            User Network Interface: Customer to SPB attach point.
   VID            VLAN ID 12 bit logical identifier after MAC header.
   VLAN           Virtual LAN: A logical network in the control plane
```

## 3. 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].

The lower case forms with an initial capital "Must", "Must Not", "Shall", "Shall Not", "Should", "Should Not", "May", and "Optional" in this document are to be interpreted in the sense defined in

[RFC2119], but are used where the normative behavior is defined in documents published by SDOs other than the IETF.

# 4. 802.1aq Overview

This section provides an overview of the behavior of [802.1aq] and is not intended to be interpreted as normative text. For the definitive behavior the reader should consult [802.1aq]. Nonetheless lower case forms with initial capitalization of the conventions in RFC2119 are used in this section to give the reader an indication of the intended normative behaviors as above.

802.1aq utilizes 802.1Q based Ethernet bridging. The filtering database (FDB) is populated as a consequence of the topology computed from the IS-IS database. For the reader unfamiliar with IEEE terminology, the definition of Ethernet behavior is almost entirely in terms of "filtering" (of broadcast traffic) rather than "forwarding" (the explicit direction of unicast traffic). This document uses the generic term "forwarding", and it has to be understood that these two terms simply represent different ways of expressing the same behaviors.

802.1aq supports multiple modes of operation depending on the type of data plane and the desired behavior. For the initial two modes of 802.1aq (SPBV and SPBM), routes are shortest path, are forward and reverse path symmetric with respect to any source / destination pair within the SPB domain, and are congruent with respect to unicast and multicast. Hence the shortest path tree (SPT) to a given node is congruent with the multicast distribution tree (MDT) from a given node. The MDT for a given VLAN is a pruned subset of the complete MDT for a given node which is identical to its SPT. Symmetry and congruency preserve packet ordering and proper fate sharing of OAM flows by the forwarding path. Such modes are fully supported by existing [802.1ag] and [Y.1731] OA&M mechanisms.

VLANs provide a natural delineation of service instances. 802.1aq supports two modes, SPB VID (SPBV) and SPB MAC (SPBM). In SPBV multiple VLANS can be used to distribute load on different shortest path trees (each computed by a different tie breaking rule) on a service basis. In SPBM service instances are delineated by I-SIDs but VLANs again can be used to distribute load on different shortest path trees.

There are two encapsulation methods supported. SPBM can be used in a PBB network implementing PBB (802.1ah [PBB]) encapsulation. SPBV can be used in PB networks implementing VLANs, PB (802.1aq [PB]) or PBB encapsulation. The two modes can co-exist simultaneously in an SPB network.

The practical design goals for SPBV and SPBM in the current 802.1aq specification are networks of size 100 nodes and 1000 nodes respectively. However since SPBV can be sparsely used in an SPB Region it can simply span a large SPB region with a small number of SPVIDs.

In SPBM and SPBV each bridge has at least one unique "known" MAC address which is advertised by IS-IS in the SYS-ID.

In the forwarding plane, SPBM uses the combination of one or more B-VIDs and "known" Backbone-MAC (B-MAC) addresses that have been advertised in IS-IS. The term Backbone simply implies an encapsulation that is often used in the backbone networks, but the encapsulation is useful in other types of networks where hiding C-MACs is useful.

The SPBM filtering database (FDB) is computed and installed for unicast and multicast MAC addresses, while the SPBV filtering database is computed and installed for unidirectional VIDs (referred to as SPVIDs), after which MAC reachability is learned (exactly as in bridged Ethernet) for unicast MACs.

Both SPBV and SPBM use source specific multicast trees. If they share the same ECT-ALGORITHM (32 bit world wide unique definition of the computation) the tree is the same SPT. For SPBV (S,G) is encoded by a source-specific VID (the SPVID) and a standard Group MAC address. For SPBM (S,G) is encoded in the destination B-MAC address as the concatenation of a 20 bit SPB wide unique nodal nickname (referred to as the SPSourceID) and the 24 bit I-SID together with the B-VID which corresponds to the ECT-ALGORITHM network wide.

802.1aq supports membership attributes which are advertised with the I-SID (SPBM) or Group Address (SPBV) that define the group. Individual members can be transmitters (T) and/or receivers (R) within the group and the multicast state is appropriately sized to these requests. Multicast group membership is possible even without transmit membership by performing head end replication to the receivers thereby eliminating transit multicast state entirely.

Some highly connected mesh networks provide for path diversity by offering multiple equal cost alternatives between nodes. Since congruency and symmetry Must be honored, a single tree may leave some links under utilized. By using different deterministic tie breakers, up to sixteen shortest paths of arbitrary diversity are possible between any pair of nodes. This distributes the traffic on a VLAN basis. SPBV and SPBM May share a single SPT with a single ECT-ALGORITHM or use any combination of the 16 ECT-ALGORITHMs. An extensible framework permits additional or alternative algorithms with other properties and parameters (e.g. ECMP, (*,G) ) to also be supported without any changes in this or the IEEE documents.

## 4.1. Multi Topology Support

SPB incorporates the multi topology features of [MT] thereby allowing multiple logical SPB instances within a single IS-IS instance.

To accomplish this, all SPB related information is either explicitly or implicitly associated with a Multi Topology Identifier (MT-ID). SPB information related to a given MT-ID thus forms a single logical SPB instance.

Since SPB has its own adjacency metrics and those metrics are also associated with an MT-ID it is not only possible to have different adjacency metrics (or infinite metrics) for SPB adjacencies, distinct from IP or other NLPIDs riding in this IS-IS instance, and also distinct from those used by other SPB instances in the same IS-IS instance.

Data plane traffic for a given MT-ID is intrinsically isolated by the VLANs assigned to the SPB instance in question. Therefore VLANs (represented by VIDs in TLVs and data plane) Must Not overlap between SPB instances (regardless of how the control planes are isolated).

The [MT] mechanism when applied to SPB allows different routing metrics and topology subsets for different classes of services.

The use of [MT] other than the default MT-ID#0 is completely OPTIONAL.

The use of [MT] to separate SPB from other NLPIDs is also OPTIONAL.

## 4.2. Data Path SPBM - Unicast

Unicast frames in SPBM are encapsulated as per 802.1ah [PBB]. A Backbone Source Address (B-SA), Backbone Destination Address (B-DA), Backbone VLAN ID (B-VID) and an I-Component Service Instance ID (I-

TAG) are used to encapsulate the Ethernet frame. The B-SA is a B-MAC associated with the ingress 802.1aq bridge, usually the "known" B-MAC of that entire bridge. The B-DA is one of the "known" B-MACs associated with the egress 802.1aq bridge. The B-VID and I-TAG are mapped based on the physical or logical UNI port (untagged, or tagged either by S-TAG or C-TAG) being bridged. Normal learning and broadcast to unknown C-MACs is applied as per [PBB] at the ingress/egress SPBs only.

Unlike [PBB] on a (*,G) tree, the B-DA forwarding on tandem nodes (NNI to NNI) is performed without learning. Instead the output of 802.1aq computations, based on the TLVs specified in this document, are used to populate the filtering data bases (FDB). The FDB entries map {B-DA, B-VID} to an outgoing interface and are only populated from the IS-IS database and computations.

The B-SA/B-VID is checked on tandem nodes against the ingress port. If the B-SA/B-VID (as a destination) entry in the FDB does not point to the port on which the packet arrived the packet is discarded. This is referred to as an Ingress Check and serves as a very powerful loop mitigation mechanism.

### 4.3. Data Path SPBM - Multicast (Head End Replication)

Head end replication is supported for instances where there is a sparse community of interest or a low likelihood of multicast traffic. Head end replication requires no Multicast state in the core. A UNI port wishing to use head end replication Must Not advertise its I-SID membership with the TX bit set but instead Must locally and dynamically construct the appropriate unicast serial replication to all the other receivers (RX) of the same I-SID.

When an unknown customer unicast or a multicast frame arrives at an SPBM User to Network Interface (UNI) port which has been configured to replicate only at the head end the packet is replicated once for each receiver, encapsulated and sent as a unicast frame. The set of receivers is determined by inspecting the IS-IS database for other SPBs that have registered interest in the same I-SID with the RX (receive) attribute set. This RX/I-SID pair is found in the SPBM Service Identifier and Unicast Address sub-TLV. The packets are encapsulated as per the SPBM Unicast forwarding above.

### 4.4. Data Path SPBM - Multicast (Tandem Replication)

Tandem replication uses the Shortest path Tree to replicate Frames only where the tree forks and there is at least one receiver on each branch. Tandem replication is bandwidth efficient but uses multicast

FDB entries (state) in core bridges which might be unnecessary if there is little multicast traffic demand. The head end replication mode is best suited for the case where there is little or no true multicast traffic for an I-SID. Tandem replication is triggered on transit nodes when the I-SID is advertised with the TX bit set.

Broadcast, unknown unicast or multicast frames arriving at an SPBM UNI port are encapsulated with a B-DA multicast address which uniquely identifies the encapsulating node (the root of the Multicast Distribution Tree) and the I-SID scoping this multicast.

This B-DA address is a well formed multicast group address (as per 802.1Q and 802.1ah) which concatenates the SPSourceID A’ with the I-SID M (written as DA=<A’,M> and uniquely identifying the (S,G) tree). This exact format is given in Figure 1 below:

```
 M L TYP
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|1|1|0|0|SPSrcMS|  SPSrc [8:15] |  SPSrc [0:7]  |  ISID [16:23] |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  ISID [8:15]  |   ISID [0:7]  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 1 SPBM Multicast Address format
(SPSrcMS represents SPSrc [16:19])

Note In Figure 1, the index numbering from less significant bit to more significant bit within a byte or field within a byte gives the wire order of the bits in the address consistent with the IETF format in the rest of this document. (The IEEE convention for number representation reverses the bits within an octet compared with IETF practice).

o  M is the multicast bit- always set to 1 for a multicast DA. (It is the lowest bit in the most significant byte.)

o  L is the local bit- always set to 1 for a SPBM constructed multicast DA.

o  TYP is the SPSourceID type. 00 is the only type supported at this time.

- o SPSRC (SPSourceID) is a 20 bit quantity that uniquely identifies a SPBM node for all B-VIDs allocated to SPBM operation. This is just the SPSourceID advertised in the SPB Instance sub-TLV. The value SPSourceID = 0 has special significance; it is advertised by an SPBM node which has been configured to assign its SPSourceID dynamically, which requires LSDB synchronization, but where the SPSourceID assignment has not yet completed.

- o I-SID is the 24 bit I component Service ID advertised in the SPBM Service Identifier TLV. It occupies the lower 24 bits of the SPBM multicast DA. The I-SID value 0xfff is reserved for SPBM control traffic(refer to the default I-SID in [802.1aq]).

This multicast address format is used as the DA on frames when they are first encapsulated at ingress to the SPBM network.  The DA is also installed into the FDBs on all SPBM nodes that are on the corresponding SPT between the source and other nodes that have registered receiver interest in the same I-SID.

Just as with unicast forwarding, the B-SA/B-VID May be used to perform an ingress check, but the SPSourceID encoded in the DA and the "drop-on-unknown" functionality of the FDB in [PBB] achieve the same effect.

The I-Component at the egress SPBM device has completely standard [PBB] behavior and therefore will:

1) learn the remote C-SA to B-SA relationship and
2) bridge the original customer frame to the set of local UNI ports that are associated with the I-SID.

## 4.5. Data Path SPBV Broadcast

When a packet for an unknown DA arrives at a SPBV UNI port VID translation (or VID encapsulation for un-tagged Frames) with the corresponding SPVID for this VLAN and ingress SPB is performed.

SPVID forwarding is simply an SPT that follows normal VLAN forwarding behavior, with the exception that the SPVID is unidirectional. As a result shared learning (SVL) is used between the forward and reverse path SPVIDs associated with the same Base-VID to allow SPBV unicast forwarding to operate in the normal reverse learning fashion.

Ingress check is done by simply verifying that the bridge to which the SPVID has been assigned is indeed "shortest path" reachable over the link over which the packet tagged with that SPVID arrived. This check is computed from the IS-IS database and is implied when the SPVID is associated with a specific incoming port.

## 4.6. Data Path SPBV Unicast

Conversely when a packet for a known DA arrives at a SPBV UNI port VID translation (or VID encapsulation for un-tagged Frames) with the corresponding SPVID for this VLAN and ingress SPB is performed.

Since the SPVID will have been configured to follow a source specific SPT and the DA is known the packet will follow the source specific path towards the destination C-MAC.

Ingress check is as per the previous SPBV section.

## 4.7. Data Path SPBV Multicast

C-DA multicast addresses May be advertised from SPBV UNI ports. These may be configured or learned through MMRP. The MMRP protocol is terminated at the edge of the SPBV network and IS-IS carries the multicast addresses. Tandem SPBV devices will check to see if they are on the SPF tree between SPBV UNI ports advertising the same C-DA multicast address, and if so will install multicast state to follow the SPBV SPF trees.

Ingress check is as per the previous two SPBV sections.

# 5. SPBM Example

Consider the following small example network shown in Figure 2. Nodes are drawn in boxes with the last nibble of their B-MAC address :1..:7, the rest of the B-MAC address nibbles are 4455-6677-00xx. Links are drawn as -- and / while the interface indexes are drawn as numbers next to the links. UNI ports are shown as <==> with the desired I-SID show at the end of the UNI ports as i1.

```
                      +----+           +----+
                      | :4 | 2 ------1 | :5 | <==> i1
                      +----+           +----+
                     1      3         3      2
                    /        \       /        \
                   1          4     3          2
                +----+        +----+          +----+
        i1 <==> | :1 | 2----1 | :2 | 2------1 | :3 | <==> i1
                +----+        +----+          +----+
                   3          6     5          3
                    \        /       \        /
                     3      2         1      2
                      +----+           +----+
                      | :6 | 1-------3 | :7 | <==> i1
                      +----+           +----+
```

Figure 2 - SPBM Example 7 node network

Using the default ECT-ALGORITHM (00-80-C2-01), which picks the equal cost path with the lowest BridgeID, this ECT-ALGORITHM is assigned to B-VID 100. When all links have the same cost, then the 1 hop shortest paths are all direct and the 2 hop shortest paths (which are of course symmetric) are as follows:

{ 1-2-3,  1-2-5, 1-2-7, 6-2-5,
  4-2-7,  4-1-6, 5-2-7, 6-2-3, 4-2-3 }

Node :1's Unicast forwarding table therefore routes toward B-MACs :7, :3 and :5 via interface/2 while its single hop paths are all direct as can be seen from its FDB given in Figure 3.

Node :1 originates multicast since it is at the head of the MDT to nodes :3, :5 and :7 and is a transmitter of I-SID 1 which nodes :3, :5 and :7 all wish to receive. Node :1 therefore produces a multicast forwarding entry who's DA contains its SPSourceID (in the example the last 20 bits of the B-MAC) and the I-SID 1 and sends to interface 2 with B-VID=100. Node :1's full unicast(U) and multicast(M) table is shown in Figure 3. Note that the IN/IF (incoming interface) field is not specified for unicast traffic and for multicast traffic has to point back to the root of the tree, unless it is the head of the tree in which case we use the convention if/OO. Since Node :1 is not transit for any multicast it only has a single entry for the root of its tree for I-SID=1.

| | IN/IF | DESTINATION ADDR | BVID | OUT/IF(s) |
|---|---|---|---|---|
| U | if/** | 4455-6677-0002 | 0100 | {if/2 } |
| U | if/** | 4455-6677-0003 | 0100 | {if/2 } |
| U | if/** | 4455-6677-0004 | 0100 | {if/1 } |
| U | if/** | 4455-6677-0005 | 0100 | {if/2 } |
| U | if/** | 4455-6677-0006 | 0100 | {if/3 } |
| U | if/** | 4455-6677-0007 | 0100 | {if/2 } |
| M | if/00 | 7300-0100-0001 | 0100 | {if/2 } |

Figure 3 - SPBM Node :1 FDB - Unicast(U) and Multicast(M)

Node :2, being at the center of the network, has direct 1 hop paths to all other nodes, therefore its unicast FDB simply sends packets with the given B-MAC/B-VID=100 to the interface directly to the addressed node. This can be seen by looking at the unicast entries (the first 6) shown in Figure 4.

| | IN/IF | DESTINATION ADDR | BVID | OUT/IF(s) |
|---|---|---|---|---|
| U | if/** | 4455-6677-0001 | 0100 | {if/1 } |
| U | if/** | 4455-6677-0003 | 0100 | {if/2 } |
| U | if/** | 4455-6677-0004 | 0100 | {if/4 } |
| U | if/** | 4455-6677-0005 | 0100 | {if/3 } |
| U | if/** | 4455-6677-0006 | 0100 | {if/6 } |
| U | if/** | 4455-6677-0007 | 0100 | {if/5 } |
| M | if/01 | 7300-0100-0001 | 0100 | {if/2,if/3,if/5 } |
| M | if/02 | 7300-0300-0001 | 0100 | {if/1 } |
| M | if/03 | 7300-0500-0001 | 0100 | {if/1,if/5 } |
| M | if/05 | 7300-0700-0001 | 0100 | {if/1,if/3 } |

Figure 4 - SPBM Node :2 FDB Unicast(U) and Multicast(M)

Node :2’s multicast is more complicated since it is a transit node for the 4 members of I-SID=1, therefore it requires 4 multicast FDB entries depending on which member it is forwarding/replicating on behalf of. For example, node :2 is on the shortest path between each of nodes {:3,:5,:7} and :1. So it must replicate from node :1 I-SID 1 out on interfaces 2, 3 and 5 (to reach nodes :3, :5 and :7). It therefore creates a multicast DA with the SPSourceID of node :1 together with I-SID=1 which it expects to receive over interface/1 and will replicate out interfaces/{2, 3 and 5}. This can be seen in the first multicast entry in Figure 4.

Note that node :2 is not on the shortest path between nodes :3 and :5 nor between nodes :3 and :7, however it still has to forward packets to node :1 from node :3 for this I-SID, which results in the second multicast forwarding entry in Figure 4. Likewise for packets originating at nodes 5 or 7, node :2 only has to replicate twice, which results in the last two multicast forwarding entries in Figure 4.

## 6. SPBV Example

Using the same example network as Figure 2, we will look at the FDBs produced for SPBV mode forwarding. Nodes :1, :5, :3 and :7 wish to transmit and receive the same multicast MAC traffic using multicast address 0300-0000-000f and at the same time require congruent and symmetric unicast forwarding. In SPBV mode the only encapsulation is the C or S-TAG and the MAC addresses SA,DA are reverse-path learned, as in traditional bridging.

```
                        +----+           +----+
                        | :4 | 2 ------1 | :5 | <==> MMAC ..:f
                        +----+           +----+
                       1      3         3      2
                      /        \       /        \
                     1          4     3          2
                  +----+        +----+          +----+
         MMAC<==> | :1 | 2----1 | :2 | 2------1 | :3 | <==> MMAC ..:f
          ..:f    +----+        +----+          +----+
                     3          6     5          3
                      \        /       \        /
                       3      2         1      2
                        +----+           +----+
                        | :6 | 1-------3 | :7 | <==> MMAC ..:f
                        +----+           +----+
```

Figure 5 - SPBV Example 7 node network

Assuming the same ECT-ALGORITHM (00-80-C2-01), which picks the equal cost path with the lowest BridgeID, this ECT-ALGORITHM is assigned to Base-VID 100, and for each node the SPVID = Base-VID + Node Id (i.e. 101, 102..107). When all links have the same cost, then the 1 hop shortest paths are all direct and the 2 hop shortest paths (which are of course symmetric) are as previously given for Figure 2.

Node :1’s SPT (Shortest Path Tree) for this ECT-ALGORITHM is therefore (described as a sequence of unidirectional paths):

{ 1->4, 1->6, 1->2->3, 1->2->5, 1->2->7 }

The FDBs therefore must have entries for the SPVID reserved for packets originating from node :1 which in this case is VID=101.

Node :2 therefore has a FDB which looks like Figure 6. In particular it takes packets from VID 101 on interface/01 and sends to nodes :3, :5 and :7 via if/2, if/3 and if/5. It does not replicate anywhere else because the other nodes :4 and :6 are reached by the SPT directly from node :1. The rest of the FDB unicast entries follow a similar pattern; recall that the shortest path between :4 and :6 is via node :1, which explains replication onto only two interfaces from if/4 and if/6. Note that the destination addresses are wild cards and shared VLAN learning (SVL) exists between these SPVIDs, because they are all associated with BASE VID = 100, which defines the VLAN being bridged.

| | IN/IF | DESTINATION ADDR | VID | OUT/IF(s) |
|---|---|---|---|---|
| U | if/01 | ************** | 0101 | {if/2,if/3,if/5 } |
| U | if/02 | ************** | 0103 | {if/1,if/4,if/6 } |
| U | if/04 | ************** | 0104 | {if/2,if/5 } |
| U | if/03 | ************** | 0105 | {if/1,if/5,if/6 } |
| U | if/06 | ************** | 0106 | {if/2,if/3 } |
| U | if/05 | ************** | 0107 | {if/1,if/3,if/4 } |

Figure 6 - SPBV Node :2 FDB unicast

Now, since nodes :5, :3, :7 and :1 are advertising membership in the same multicast group address :f, Node 2 requires additional entries to replicate just to these specific nodes for the given multicast group address. These additional multicast entries are given below in Figure 7.

| | IN/IF | DESTINATION ADDR | VID | OUT/IF(s) |
|---|---|---|---|---|
| M | if/01 | 0300-0000-000f | 0101 | {if/2,if/3,if/5 } |
| M | if/02 | 0300-0000-000f | 0103 | {if/1 } |
| M | if/03 | 0300-0000-000f | 0105 | {if/1,if/5 } |
| M | if/05 | 0300-0000-000f | 0107 | {if/1,if/3 } |

Figure 7 - SPBV Node :2 FDB Multicast(M)

## 7. SPB Supported Adjacency types

IS-IS for SPB currently only supports P2P adjacencies. Other link types are for future study. As a result pseudonodes and links to/from pseudonodes are not considered as part of the IS-IS SPF computations and will be avoided if present in the physical topology. Other NLPIDs MAY of course use them as per normal.

IS-IS for SPB Must use the IS-IS Three-Way handshake for IS-IS Point-to-Point Adjacencies described in RFC 5303.

## 8. SPB IS-IS adjacency addressing

The default behavior of 802.1aq is to use the normal IS-IS Ethernet multicast addresses for IS-IS.

There are however additional Ethernet multicast addresses that have been assigned for 802.1aq for special use cases. These do not in anyway change the state machinery or packet formats of IS-IS but simply recommend and reserve different multicast addresses. Refer to [802.1aq] for additional details.

## 9. IS-IS Area Address and SYSID

A stand-alone implementation (supporting ONLY the single NLPID=0xC1) of SPB Must use an IS-IS area address value of 0 and the SYSID Must be the well known MAC address of the SPB device.

Non stand-alone implementations (supporting other NLPIDs) MUST use the normal IS-IS rules for the establishment of a level 1 domain (i.e. multiple area addresses are allowed but where immediate adjacencies share a common area address). Level 2 operations of course place no such restriction on adjacent area addresses.

## 10. Level 1/2 Adjacency

SPBV and SPBM will operate either within an IS-IS level 1, or an ISIS level 2. As a result, the TLVs specified here MAY propagate either in level 1 or level 2 LSPs. IS-IS SPB implementations Must support level 1 and May support level 2 operations. Hierarchical SPB is for further study therefore these TLV’s Should Not be leaked between level 1 and level 2.

## 11. Shortest Path Default Tie Breaking

(ECT-ALGORITHM = 00-80-C2-01)

Two mechanisms are used to ensure symmetry and determinism in the shortest path calculations.

The first mechanism addresses the problem when different ends (nodes) of an adjacency advertise different values for the SPB-LINK-METRIC. To solve this SPB shortest path calculations Must use the maximum value of the two node's advertised SPB-LINK-METRICs when accumulating and minimizing the (sub)path costs.

The second mechanism addresses the problem when two equal sums of link metrics (sub)paths are found. To solve this, the (sub)path with the fewest hops between the fork/join points Must win the tie. However, if both (sub)paths have the same number of hops between the fork and join points then the default tie breaking Must pick the path traversing the intermediate node with the lower BridgeID. The BridgeID is an 8 byte quantity whose upper 2 bytes are the node's BridgePriority and the lower 6 bytes are the node's SYSID.

For example, consider the network in Figure 2 when a shortest path computation is being done from node :1. Upon reaching node :7 two competing sub-paths fork at node :1 and join at node :7. The first via :2 and the second via :6. Assuming that all the nodes advertise a Bridge Priority of 0, the default tie breaking rule causes the path traversing node :2 to be selected since it has a lower BridgeID {0...:2} than node :6 {0...:6}. Note that the operator may cause the tie breaking logic to pick the alternate path by raising the Bridge Priority of node :2 above that of node :6.

The above algorithm guarantees symmetric and deterministic results in addition to having the critical property of transitivity (shortest path is made up of sub-shortest paths).

## 12. Shortest Path ECT

(ECT-ALGORITHMs = 00-80-C2-01 .. 00-80-C2-10)
To create diversity in routing SPB defines 16 variations on the above default tie breaking algorithm, these have world wide unique designations 00-80-C2-01 through 00-80-C2-10. These designations consist of the IEEE 802.1 OUI value 00-80-C2 concatenated with indexes 0X01..0X10. These individual algorithms are implemented by selecting the (sub) path with the lowest value of:

```
        XOR BYTE BY BYTE(ECT-MASK{ECT-ALGORITHM.index},BridgeID)

   Where:

        ECT-MASK{17} = { 0x00, 0x00, 0xFF, 0x88,
                         0x77, 0x44, 0x33, 0xCC,
                         0xBB, 0x22, 0x11, 0x66,
                         0x55, 0xAA, 0x99, 0xDD,
                         0xEE };

        XOR BYTE BY BYTE  - XORs BridgeID bytes with ECT-MASK
```

ECT-MASK{1} since it xor’s with all 0’s is just the same as the default algorithm described above 00-80-C2-01, while ECT-MASK{0x02} since it xor’s with a mask of all 1’s will invert the BridgeID essentially picking the path traversing the largest Bridge ID. The other ECT-MASKs produce diverse alternatives. In all cases the BridgePriority, since it is the most significant part of the BridgeID permits overriding the SYSID as the selection criteria and gives the operator a degree of control on the chosen ECT paths.

To support many other tie breaking mechanisms in the future two opaque ECT TLV’s are defined which may be used to provide parameters to ECT-ALGORITHMS outside of the currently defined space.

ECT-ALGORITHMS are mapped to VIDs and then services can be assigned to those VIDs. This permits a degree of traffic engineering since service assignment to VID is consistent end to end through the network.

## 13. Hello (IIH) protocol extensions

IEEE 802.1aq can run in parallel with other Network Layer Protocols such as IPV4 and IPV6, therefore failure for two SPB nodes to establish an adjacency MUST NOT cause rejection of an adjacency for the purposes of other Network Layer Protocols.

IEEE 802.1aq has been assigned the NLPID value 0xC1 [NLPID] which MUST be used by shortest path bridges (SPBs) to indicate their ability to run 802.1aq.  This is done by including this NLPID value in the IS-IS IIH PDU Protocols Supported TLV (type 129). 802.1aq frames MUST only flow on adjacencies that advertise this NLPID in both directions of the IIH PDUs. 802.1aq computations MUST consider an adjacency that has not advertised 0xC1 NLPID in both directions as non-existent (infinite link metric) and MUST ignore any IIH SPB TLV’s they receive over such adjacencies.

IEEE 802.1aq augments the normal IIH PDU with three new TLV’s which like all other SPB TLVs travel within multi topology [MT] TLVs, therefore allowing multiple logical instances of SPB within a single IS-IS protocol instance.

Since SPB can use many VIDs and Must agree on which VIDs are used for which purposes, the IIH PDU’s carry a digest of all the used VIDs (on the NNI’s) referred to as the SPB-MCID TLV which uses a common and compact encoding taken reused from 802.1Q.

SPB neighbors May support a mechanism to verify that the contents of their topology databases are synchronized (for the purposes of loop prevention). This is done by exchanging a digest of SPB topology information (computed over all MTIDS) and taking specific actions on forwarding entries when the digests indicate a mismatch in topology. This digest is carried in the Optional SPB Digest sub-TLV.

Finally SPB needs to know which SPT sets (defined by ECT-ALGORITHMS) are being used by which VIDs, and this is carried in the Base VLAN Identifiers sub-TLV.

## 13.1. SPB MCID sub-TLV

This sub-TLV is added to an IIH PDU to indicate the digest for the Multiple spanning tree configuration a.k.a MCID. This TLV is a digest of local configuration of which VIDs are running which protocols. (The information is not to the level of a specific algorithm in the case of SPB). This information Must be the same on all bridges in the SPT Region controlled by an IS-IS instance. The data used to generate the MCID is populated by configuration and is a digest of the VIDs allocated to various protocols. Two MCIDs are carried to allow non disruptive transitions between configurations when the changes are non-critical.

```
    0                   1                   2                   3
    0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1

   +-+-+-+-+-+-+-+-+
   |Type=SPB-MCID  | = 6
   +-+-+-+-+-+-+-+-+
   |   Length      |    (1 byte)
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+...+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                          MCID (51 Bytes)                        |
   |                          ...............                        |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+...+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                    Aux   MCID (51 Bytes)                        |
   |                          ...............                        |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+...+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

- Type: sub-TLV Type = 6 (Pending IANA).

- Length: The size of the value defined below (102).

- MCID (51-bytes) The complete MCID defined in IEEE 802.1Q which identifies an SPT Region on the basis of matching assignments of VIDs to control regimes (xSTP, SPBV, SPBM, etc). Briefly, the MCID consists of a 1 byte format selector, a 32 byte configuration name, a 2 byte revision level and finally a 16 byte signature of type HMAC-MD5 over an array of 4096 elements that contain identifiers of the use of the corresponding VID. Refer to section 13.8 of [802.1aq] for the exact format and procedure. Note that the use of the VID does not include specification of a specific SPB ECT-ALGORITHM, rather it is coarser grain.

- Aux MCID (51-bytes) The complete MCID defined in IEEE 802.1Q which identifies an SPT Region.  The aux MCID allows SPT Regions to be migrated by the allocation of new VLAN to FDB Mappings without interruption to existing traffic.

The SPB MCID sub-TLV is carried within the MT-Port-Cap TLV [LAYER2] with the MT-ID value of 0, which in turn is carried in an IIH PDU.

## 13.2. SPB Digest sub-TLV

This sub-TLV is Optionally added to an IIH PDU to indicate the current SPB topology digest value. It is always carried in an MT-Port-Cap TLV [LAYER2] with an MT-ID value of 0. This information should settle to be the same on all bridges in an unchanging topology. Matching digests indicate (with extremely high probability) that the topology view between two SPBs is synchronized, and is used to control the updating of forwarding information.  The SPB Agreement Digest is computed based on contributions derived from the current topologies of all SPB MT instances, and is designed to change when significant topology changes occur within any SPB instance.

During the propagation of LSPs the Agreement Digest may vary between neighbors until the key topology information in the LSPs are common. The digest is therefore a summarized means of determining agreement between nodes on database commonality, and hence infer agreement on the distance to all multicast roots. When present it is used for loop prevention as follows:  For each shortest path tree where it has been determined the distance to the root has changed, "unsafe" multicast forwarding is blocked until the exchanged Agreement Digests match while "safe" multicast forwarding is allowed to continue despite the disagreement in digests and hence topology views. [802.1aq] section 28.2 defines in detail what constitutes "safe" vs. "unsafe".

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1

+-+-+-+-+-+-+-+-+
|Type=SPB-Digest| = 7
+-+-+-+-+-+-+-+-+
|   Length      | (1 byte)
+-----+-+---+---+
| Res |V| A | D | (1 byte)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+...+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|               Agreement Digest (Length - 1)                     |
|                            ...............                      |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+...+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

o  Type: sub-TLV Type = 7 (Pending IANA).

o  Length: The size of the value.

o  V - agreed digest valid bit. See [802.1aq] Sec 28.2.

- o A (2 bits) The Agreement Number 0-3 which aligns with BPDUs Agreement Number concept [802.1aq]. When the Agreement Digest for this node changes this number is incremented. The node then checks for Agreement Digest match (as below). The new local Agreement Number and the updated local Discarded Agreement Number are then transmitted with the new Agreement Digest to the node’s neighbors in the hello PDU. Once an Agreement Number has been sent it is considered outstanding until a matching or more recent Discarded Agreement Number is received from the neighbor.

- o D (2 bits) The Discarded Agreement Number 0-3 which aligns with BPDUs Agreement Number concept. When an Agreement Digest is received from a neighbor, this number is set to the received Agreement Number, to signify that this node has received this new agreement and discarded any previous ones. The node then checks whether the local and received Agreement Digests match. If they do, this node then sets :

  the local Discarded Agreement Number = received Agreement Number + 1

  If the Agreement Digests match, AND received Discarded Agreement Number == local Agreement Number + N (N = 0 || 1)

  then the node has a topology matched to its neighbor.

  Whenever the local Discarded Agreement Number relating to a neighbor changes, the local Agreement Digest, Agreement Number, and Discarded Agreement Number are transmitted.

- o Agreement Digest. This digest is used to determine when SPB is synchronized between neighbors for all SPB instances. The agreement digest is a hash computed over the set of all SPB adjacencies in all SPB instances. In other words, the digest includes all VIDs and all adjacencies for all MT instances of SPB (but not other network layer protocols). This reflects the fact that all SPB nodes in a region Must have identical VID allocations (see 13.1), and so all SPB instances will contain the same set of nodes. The size and exact procedure for computing the Agreement Digest is defined in section 28.2 of [802.1aq].

The SPB Digest sub-TLV is carried within the MT-Port-Cap TLV [LAYER2] (with the MT-ID value 0) which in turn is carried in an IIH PDU.

When supported, this sub-TLV MUST be carried on every IIH between SPB neighbors, not just when a Digest changes.

When one peer supports this TLV and the other does not, loop prevention by digest agreement Must Not be done by either side.

13.3. SPB Base VLAN-Identifiers sub-TLV

This sub-TLV is added to an IIH PDU to indicate the mappings between ECT algorithms and Base-VIDs (and by implication the VID(s) used on the forwarding path for each SPT Set for a VLAN identified by a Base VID) that are in use. Under stable operational conditions, this information should be the same on all bridges in the topology identified by the MT-Port-Cap TLV [LAYER2] it is being carried within.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1

+-+-+-+-+-+-+-+-+
|Type= SPB-B-VID| = 8
+-+-+-+-+-+-+-+-+
|   Length      |    (1 byte)
+-+-+-+-+-+-+-+-+-----------------------------------------------+
|      ECT - VID Tuple (1)  (6 bytes)                           |
+---------------------------------+-----------------------------+
|      ...                        | ECT-VID Tuple(2) (6 bytes)  |
+---------------------------------+-----------------------------+
|                          .....                                |
+---------------------------------------------------------------+
|                          .....                                |
|                          .....                                |
+---------------------------------------------------------------+
```

- o Type: sub-TLV Type = 8 (Pending IANA).

- o Length: The size of the value is ECT-VID Tuples*6 bytes. Each 6-byte part of the ECT-VID tuple is formatted as follows:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1

+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     ECT - Algorithm (32 bits)                 |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Base VID (12 bits)    |U|M|RES|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

- o ECT-ALGORITHM (4 bytes) The ECT-ALGORITHM is advertised when the bridge supports a given ECT-ALGORITHM (by OUI/Index) on a given Base-VID. There are 17 predefined IEEE algorithms for SPB with index values 0X00..0X10 occupying the low 8 bits and the IEEE OUI=00-80-C2 occupying the top 24 bits of the ECT-ALGORITHM.

- o Base-VID (12-bits) The Base-VID that is associated with the SPT Set.

- o Use-Flag (1-bit) The Use-flag is set if this bridge, or any bridge in the LSDB is currently using this ECT-ALGORITHM and Base-VID. Remote usage is discovered by inspection of the U-Bit in the SPB Instance sub-TLV of other SPB bridges (see section 14.1)

- o M-Bit (1-bit) The M-bit indicates if this Base-VID operates in SPBM (M = 1) or SPBV (M = 0) mode.

The SPB Base VLAN-Identifier sub-TLV is carried within the MT-Port-Cap TLV [LAYER2] which in turn is carried in an IIH PDU.

## 14. Node information extensions

All SPB nodal information extensions travel within a new multi topology capability TLV MT-Capability (type = 144).

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1

+-+-+-+-+-+-+-+-+
|Type = MT-CAP  | = 144
+-+-+-+-+-+-+-+-+
|   Length      |     (1 byte)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|O R R R|       MT ID           | (2 bytes)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
  (sub-TLVs ... )
```

The format of this TLV is identical in its first 2 bytes to all current MT TLV’s and carries the MT ID as defined in [MT].

The O (overload) bit carried in bit 16 has the same semantics as specified in [MT], but in the context of SPB adjacencies only.

## 14.1. SPB Instance sub-TLV

The SPB Instance sub-TLV gives the SPSourceID for this node/topology instance. This is the 20 bit value that is used in the formation of multicast DA addresses for frames originating from this node/instance. The SPSourceID occupies the upper 20 bits of the multicast DA together with 4 other bits (see the SPBM 802.1ah multicast DA address format section). This sub-TLV MUST be carried within the MT-Capability TLV in the fragment ZERO LSP. If there is an additional SPB instance it MUST be declared under a separate MT-Topology and also carried in the fragment ZERO LSP.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1

+-+-+-+-+-+-+-+-+-+
|Type = SPB-Inst| = 1
+-+-+-+-+-+-+-+-+-+
|   Length      |     (1 byte)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|               CIST Root Identifier  (4 bytes)                 |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|               CIST Root Identifier (cont)  (4 bytes)          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|           CIST External ROOT Path Cost     (4 bytes)          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|        Bridge Priority        |          (2 bytes)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|R R R R R R R R R R R|V|              SPSourceID               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Num of Trees  |       (1 byte)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                  VLAN-ID (1) Tuples    (8 bytes)              |
|                  ..........................                   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
                   ..........................
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                  VLAN-ID (N) Tuples    (8 bytes)              |
|                  ..........................                   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

where VLAN-ID tuples have the format as:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+
|U|M|A|  Res    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     ECT - Algorithm (32 bits)                 |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Base-VID (12 bits)    |   SPVID (12 bits)     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

o  Type: sub-TLV Type 1 (Pending IANA).

o  Length: Total number of bytes contained in the value field.

o  CIST Root Identifier (64-bits)The CIST Root Identifier is for SPB interworking with RSTP and MSTP at SPT Region Boundaries.  This is an imported value from a Spanning tree.

o  CIST External Root Path Cost (32-bits) The CIST External Root Path Cost is the cost to root, derived from the spanning tree algorithm.

o  Bridge Priority (16-bits) Bridge priority is the 16 bits that together with the 6 bytes of the System ID form the Bridge Identifier. This is configured exactly as specified in IEEE802 [802.1D]. This allows SPB to build a compatible Spanning tree using link state by combining the Bridge Priority and the System ID to form the 8 byte Bridge Identifier.  The 8 byte Bridge Identifier is also the input to the 16 pre-defined ECT tie breaker algorithms.

o  V bit (1-Bit) The V bit (SPBM) indicates this SPSourceID is auto allocated(27.11).  If the V bit is clear the SPSourceID has been configured and Must be unique.  Allocation of SPSourceID is defined in IEEE [802.1aq].  Bridges running SPBM will allocate an SPSourceID if they are not configured with an explicit SPSourceID. The V Bit allows neighbor bridges to determine if the auto allocation was enabled.  In the rare chance of a collision of SPsourceID allocation, the bridge with the highest priority Bridge Identifier will win conflicts and the lower priority Bridge will be re-allocated or if the lower priority Bridge is configured it will not be allowed to join the SPT Region.

- o The SPSourceID is a 20 bit value used to construct multicast DA's as described below for multicast frames originating from the origin (SPB node) of the link state packet (LSP) that contains this TLV. More details are in IEEE [802.1aq].

- o Number of Trees (8-bits) The Number of Trees is set to the number of [ECT-ALGORITHM, Base-VID plus flags] tuples that follow. Each ECT-ALGORITHM has a Base-VID, an SPVID and flags described below. This Must contain at least the one ECT-ALGORITMM (00-80-C2-01).

Each VID Tuple consists of:

- o U-Bit (1-bit) The U-bit is set if this bridge is currently using this ECT-ALGORITHM for I-SIDs it sources or sinks. This is a strictly local indication; the semantics differ from the Use-flag found in the Hello, which will set the Use-Flag if it sees other nodal U-bits are set OR it sources or sinks itself.

- o M-Bit (1-bit) The M-bit indicates if this is SPBM or SPBV mode. When cleared the mode is SPBV and when set the mode is SPBM.

- o A bit, The A bit (SPB) when set declares this is an SPVID with auto allocation. The VID allocation logic details are in IEEE [802.1aq]. Since SPVIDs are allocated from a small pool of 12 bit resources the chances of collision are high. To minimize collisions during auto allocation LSPs are initially advertised with the originating bridge setting the SPVID to 0. Only after learning the other bridges' SPVID allocations does this bridge re-advertise this sub-TLV with a non-zero SPVID. This will minimize but not eliminate the chance of a clash. In the event of a clash, the highest Bridge Identifier is used to select the winner, while the loser(s) with lower Bridge Identifier(s) Must withdraw their SPVID allocation(s), and select an alternative candidate for another trial. SPVID May also be configured. When the A bit is set to not specify auto allocation and the SPVID is set to 0 this SPBV bridge is used for transit only within the SPB region. If a port is configured with the BASE-VID as a neighbor using RSTP or MSTP the bridge will act as an ingress filter for that VID.

- o ECT-ALGORITHM (4-bytes) ECT-ALGORITHM is advertised when the bridge supports a given ECT-ALGORITHM (by OUI/Index) on a given VID. This declaration Must match the declaration in the Hello PDU originating from the same bridge. The ECT-ALGORITHM and BASE-VID Must match what is generated in the IIHs of the same node. The ECT-ALGORITHM, BASE-VID tuples can come in any order however. There are currently 17 world wide unique 802.1aq defined ECT-ALGORITHMS given by values 00-80-C2-00 through 00-80-C2-10.

- o Base VID (12-bits) The Base-VID that associated the SPT Set via the ECT-ALGORITHM.

- o SPVID (12-bits) The SPVID is the Shortest Path VID assigned for the Base-VID to this node when using SPBV mode. It is not defined for SPBM Mode and Must be 0 for SPBM mode B-VIDs.

### 14.1.1. SPB Instance Opaque ECT-ALGORITHM sub-TLV

There are multiple ECT algorithms defined for SPB, however for the future additional algorithms may be defined including but not limited to ECMP / hash based behaviors and (*,G) multicast trees. These algorithms will use this Optional TLV to define new algorithm parametric data. For tie breaking parameters there are two broad classes of algorithm, one which uses nodal data to break ties and one which uses link data to break ties, this TLV is used to associate opaque tie breaking data with a node. This sub-TLV, when present, MUST be carried within the MT-Capability TLV (along with a valid SPB Instance sub-TLV). Multiple copies of this sub-TLV MAY be carried for different ECT-ALGORITHMs relating to this node.

There are of course many other uses of this opaque data which have yet to be defined.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1

+-+-+-+-+-+-+-+-+
|Type=SPB-I-OALG| = 2
+-+-+-+-+-+-+-+-+
|   Length      |     (1 byte)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                  Opaque ECT Algorithm    (4 bytes)            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                  Opaque ECT Information (variable)            |
|                   .......................                     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

o  Type: sub-TLV Type 2 (Pending IANA).

o  Length: Total number of bytes contained in the value field.

o  ECT-ALGORITHM: ECT-ALGORITHM is advertised when the bridge supports a given ECT-ALGORITHM (by OUI/Index) on a given VID.

o  ECT Information: ECT-ALGORITHM Information of variable length which SHOULD be in sub-TLV format with an IANA numbering space where appropriate.

## 15. Adjacency information extensions

### 15.1. SPB Link Metric sub-TLV

The SPB Link Metric sub-TLV (type = 12) occurs within the Multi Topology Intermediate System TLV (type 222) or within the Extended IS Reachability TLV (type 22).  If this sub TLV is not present for an ISIS adjacency then that adjacency Must not carry SPB traffic for the given topology instance.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1

+-+-+-+-+-+-+-+-+
|Type=SPB-Metric| = 12
+-+-+-+-+-+-+-+-+
|   Length      |     (1 byte)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|       SPB-LINK-METRIC                         |   (3 bytes)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Num of ports    |     (1 byte)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     Port  Identifier          |   ( 2 bytes)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

o  Type: sub-TLV Type 12 (Pending IANA).

o  Length: Total number of bytes contained in the value field.

o SPB-LINK-METRIC indicates the administrative cost or weight of using this link as a 24 bit unsigned number. This metric applies to the use of this link for SPB traffic only. Smaller numbers indicate lower weights and are more likely to carry SPB traffic. Only one metric is allowed per SPB instance per link. If multiple metrics are required multiple SPB instances Must be used, either within IS-IS or within several independent IS-IS instances. If this metric is different at each end of a link, the maximum of the two values Must be used in all SPB calculations for the weight of this link. The maximum SPB-LINK-METRIC value 2^24 - 1 has a special significance; this value indicates that although the IS-IS adjacency has formed, incompatible values have been detected in parameters configured within SPB itself for example different regions, and the link Must Not be used for carrying SPB traffic. Full details are found in [802.1aq].

o Num of Ports is the number of ports associated with this link.

o Port Identifier is the standard IEEE port identifier used to build a spanning tree associated with this link.

15.1.1. SPB Adjacency Opaque ECT-ALGORITHM sub-TLV

There are multiple ECT algorithms defined for SPB, however for the future additional algorithms may be defined. The SPB Adjacency Opaque ECT-ALGORITHM sub-TLV occurs within the Multi Topology Intermediate System TLV (type 222) or the Extended IS Reachability TLV (type 22). Multiple copies of this sub-TLV MAY be carried for different ECT-ALGORITHMs related to this adjacency.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1

+-+-+-+-+-+-+-+-+
|Type=SPB-A-OALG| = 13
+-+-+-+-+-+-+-+-+
|   Length      |     (1 byte)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                  Opaque ECT Algorithm    (4 bytes)            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                  Opaque ECT Information (variable)            |
|                  .........................                    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

o Type: sub-TLV Type = 13 (PENDING IANA).

o Length: Total number of bytes contained in the value field.

o  ECT-ALGORITHM: ECT-ALGORITHM is advertised when the bridge supports a given ECT-ALGORITHM (by OUI/Index) on a given VID.

o  ECT Information: ECT-ALGORITHM Information of variable length in sub-TLV format using new IANA type values as appropriate.

## 16. Service information extensions

### 16.1. SPBM Service Identifier and Unicast Address sub-TLV

The SPBM Service Identifier and Unicast Address sub-TLV (type=3) is used to introduce service group membership on the originating node and/or to advertise an additional B-MAC unicast address present on, or reachable by the node.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1

+-+-+-+-+-+-+-+-+
|Type = SPBM-SI | = 3
+-+-+-+-+-+-+-+-+
|   Length      |     (1 byte)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                       B-MAC ADDRESS                           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|    B-MAC ADDRESS  (6 bytes)   |  Res. |   Base-VID (12 bits)  |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|T|R| Reserved  |                  ISID  #1                     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|T|R| Reserved  |                  ISID  #2                     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
                          .................
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|T|R| Reserved  |                  ISID  #n                     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

o  Type: sub-TLV Type = 3 (Pending IANA)

o  Length: Total number of bytes contained in the value field.

- o B-MAC ADDRESS is a unicast address of this node. It may be either the single nodal address, or may address a port or any other level of granularity relative to the node. In the case where the node only has one B-MAC address this Should be the same as the SYS-ID of the node. To add multiple B-MACs this TLV MUST be repeated per additional B-MAC.

- o Base VID (12-bits) The Base-VID associated with the B-BMAC this allows the linkage to the ECT-Algorithm and SPT Set defined in the SPB Instance sub-TLV.

- o ISID #1 .. #N are 24 bit service group membership identifiers. If two nodes have an I-SID in common, intermediate nodes on the unique shortest path between them will create forwarding state for the related B-MAC addresses and will also construct multicast forwarding state using the I-SID and the node's SPSourceID to construct a multicast DA as described in IEEE 802.1aq LSB. Each I-SID has a Transmit(T) and Receive(R) bit which indicates if the membership is as a Transmitter/Receiver or both (with both bits set). In the case where the Transmit(T) and Receive(R) bits are both zero, the I-SID instance is ignored for the purposes of distributed multicast computation, but the unicast B-MAC address Must be processed and installed at nodes providing transit to that address. If more I-SIDs are associated with a particular B-MAC than can fit in a single sub-TLV, this sub-TLV can be repeated with the same B-MAC but with different I-SID values.

- o Note when the T bit is not set an SPB May still multicast to all the other receive members of this I-SID (those advertising with their R bits set), by configuring edge replication and serial unicast to each member locally.

The SPBM Service Identifier sub-TLV, when present, MUST be carried within the MT Capability TLV and can occur multiple times in any LSP fragment.

## 16.2. SPBV Mac Address sub-TLV

The SPBV MAC Address (SPBV-MAC-ADDR) sub-TLV is IS-IS sub-TLV type 4 (PENDING IANA). It Should be used for advertisement of Group MAC Addresses in SPBV mode. Unicast MAC addresses will normally be distributed by reverse path learning, but carrying them in this TLV is not precluded. It has the following format :

```
    0                   1                   2                   3
    0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1

   +-+-+-+-+-+-+-+-+
   | Type=SPBV-ADDR|   = 4           (1 byte)
   +-+-+-+-+-+-+-+-+
   |   Length      |                 (1 byte)
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |R|R|S-R|       SPVID           |  (2 bytes)
   +-+-+-+-+-+-+-+-+-+-+-+-+-+...+-+-+-+-+-+-+-+-+-+-+-+-+
   |T|R| Reserved  |      MAC 1 Address              |  (1+6 bytes)
   +-+-+-+-+-+-+-+-+-+-+-+-+-+...+-+-+-+-+-+-+-+-+-+-+-+-+
                             ...
   +-+-+-+-+-+-+-+-+-+-+-+-+-+...+-+-+-+-+-+-+-+-+-+-+-+-+
   |T|R| Reserved  |      MAC N Address              |  (1+6 bytes)
   +-+-+-+-+-+-+-+-+-+-+-+-+-+...+-+-+-+-+-+-+-+-+-+-+-+-+
```

o  Type: sub-TLV Type, set to 4.

o  Length: Total number of bytes contained in the value field.  The number of MAC address associated with the SPVID is computed by (Length - 2)/7.

o  S-R bits (2-bits) The SR bits are the service requirement parameter from MMRP.  The service requirement parameters have the value 0 (Forward all Groups) and 1 (Forward All Unregistered Groups) defined.  However this attribute May also be missing.  So the SR bits are defined as 0 not declared, 1 Forward all Groups and 2 Forward All Unregistered Groups.  The two 'R' reserved bits immediately preceding these SR bits Shall be set to zero when originating this sub-TLV and Shall be ignored on receipt.

o  SPVID (12-bits) The SPVID and by association Base-VID and the ECT-ALGORITHM and SPT Set that the MAC addresses defined below will use. If the SPVID is not allocated the SPVID Value is 0. Note that if the ECT-Algorithm in use is Spanning Tree Algorithm this value Must be populated with the Base-VID and the MAC Must be populated.

- T Bit (1-bit) This is the Transmit allowed Bit for a following group MAC address. This is an indication that the Group MAC Address in the context of the SPVID of the bridge advertising this Group MAC Must be installed in the FDB of transit bridges, when the bridge computing the trees is on the corresponding ECT-ALGORITHM shortest path between the bridge advertising this MAC with the T bit set and any receiver of this Group MAC Address. A bridge that does not advertise this bit set for a MAC Address Must Not cause multicast forwarding state to be installed on other transit bridges in the network for traffic originating from that bridge.

- R Bit (1-bit) This is the Receive allowed Bit for the following MAC Address. This is an indication that MAC Addresses as receiver Must be populated and installed when the bridge computing the trees lies on the corresponding shortest path for this ECT-ALGORITHM between this receiver and any transmitter to this MAC Address. An entry that does not have this bit set for a Group MAC Address is prevented from receiving on this Group MAC Address because transit bridges Must Not install multicast forwarding state towards it in their FDBs.

- MAC Address (48-bits) The MAC address declares this bridge as part of the multicast interest for this destination MAC address. Multicast trees can be efficiently constructed for destination by populating FDB entries for the subset of the shortest path tree that connects the bridges supporting the MAC address. This replaces the function of MMRP for SPTs. The T and R bits above have meaning as specified above.

The SPBV-MAC-ADDR sub-TLV, when present, MUST be carried within the MT-Capability TLV and can occur multiple times in any LSP fragment.

## 17. Security Considerations

This document adds no additional security risks to IS-IS, nor does it provide any additional security for IS-IS when used in a configured environment or a single operator domain such as a Data Center.

However this protocol may be used in a zero configuration environment. Zero configuration may apply to the automatic detection and formation of an IS-IS adjacency (forming an NNI port). Likewise zero configuration may apply to the automatic detection of VLAN tagged traffic and the formation of a UNI port, with resultant ISID advertisements.

If zero configuration methods are used to auto configure NNIs or UNIs there are intrinsic security concerns that should be mitigated with authentication procedures for the above cases. Such procedures are beyond the scope of this document, and are yet to be defined.

In addition, this protocol can create significant amounts of multicast state when an ISID is advertised with the TX bit set. Extra care should be taken to ensure that this cannot be used in a Denial of Service attack [RFC4732] in a zero configuration environment.

## 18. IANA Considerations

Note that the NLPID value 0xC1 [NLPID] used in the IIH PDUs has already been assigned by IANA for the purpose of 802.1aq therefore no further action is required for this code point.

Since 802.1aq operates within the IS-IS Multi Topology framework every sub-TLV MUST occur in the context of the proper MT TLV (with the exception of the SPB Link Metric sub-TLV which MAY travel in TLV 22 where its MT-ID is unspecified but implied to be 0). There are three Multi Topology TLV's in which 802.1aq requests allocation of sub-TLV's. These are the MT-Port-Cap TLV [LAYER2] used in the IIH, the MT-Capability TLV (new) used within the LSP and finally the MT-Intermediate-System TLV [MT] used to contain adjacency information within the LSP.

This document creates the following TLVs & sub-TLV's within the IIH and LSP PDUs MT TLV's as described below. The '*' indicates IANA action is required. Other entries are shown to provide context only. A '?' next to a number indicates a requested but of course not necessarily the final assigned value.

The MT-Capability TLV is the only TLV requiring a new sub-registry. Type value 144 (TBD) is requested, with a starting sub-TLV value of 1, and managed by Expert Review.

```
      +-----+----+-----------------+--------+------+-------------+
      | PDU |TLV | SUB-TLV         | TYPE   | TYPE | #OCCURRENCE |
      +-----+----+-----------------+--------+------+-------------+
        IIH
             MT-Port-Cap              147?
   *               SPB-MCID                   6?     1
   *               SPB-Digest                 7?     >=0
   *               SPB-B-VID                  8?     1

        LSP
   *         MT-Capability            144?
   *               SPB-Inst                   1?     1
   *               SPB-I-OALG                 2?     >=0
   *               SPBM-SI                    3?     >=0
   *               SPBV-ADDR                  4?     >=0

             MT-Intermediate-System    222
          or Extended IS Reachability   22
   *               SPB-Metric                12?     1
   *               SPB-A-OALG                13?     >=0
```

## 19. References

### 19.1. Normative References

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

[IS-IS] ISO/IEC 10589:2002, Second Edition, "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)", 2002.

[MT] M-ISIS: Multi Topology (MT) Routing in Intermediate System to Intermediate Systems (IS-ISs), RFC 5120, February 2008.

[802.1aq] "Standard for Local and Metropolitan Area Networks / Virtual Bridged Local Area Networks / Amendment: Shortest Path Bridging, IEEE P802.1aq  Draft 3.6", 2011.

[NLPID] www.ietf.org/id/draft-eastlake-nlpid-iana-considerations-04.txt

[LAYER2] www.ietf.org/id/draft-ietf-isis-layer2-09.txt.

## 19.2. Informative References

[MMRP] "Standard for Local and Metropolitan Area Networks Virtual Bridged Local Area Networks - Amendment 07: Multiple Registration Protocol", IEEE STD 802.1ak, 2007

[PB] "Standard for Local and Metropolitan Area Networks / Virtual Bridged Local Area Networks / Amendment 4: Provider Bridges, IEEE STD 802.1ad", 2005.

[PBB] "Standard for Local and Metropolitan Area Networks / Virtual Bridged Local Area Networks / Amendment 7: Provider Backbone Bridges, IEEE STD 802.1ah", 2008.

[802.1ag] "Standard for Local and Metropolitan Area Networks / Virtual Bridged Local Area Networks / Amendment 5: Connectivity Fault Management", IEEE STD 802.1ag, 2007

[Y.1731] ITU-T Y.1731 (2006), "OAM Functions and Mechanisms for Ethernet based networks"

[RFC4732] Handley, M., Ed, "Internet Denial-of-Service Considerations", RFC 4732, November 2006.

## 20. Acknowledgments

The authors would like to thank Ayan Banerjee, Mick Seaman, Janos Farkas, Les Ginsberg, Stewart Bryant , Donald Eastlake, Matthew Bocci and Mike Shand for contributions and/or detailed review.

This document was prepared using 2-Word-v2.0.template.dot.

## 21. Author’s Addresses

Don Fedyk
Alcatel-Lucent
Groton, MA, 01450, USA
Email: Donald.Fedyk@alcatel-lucent.com

Peter Ashwood-Smith
Huawei Technologies Canada Ltd,
303 Terry Fox Drive, Suite 400
Kanata, Ontario, K2K 3J1, CANADA
Email: Peter.AshwoodSmith@huawei.com

Dave Allan
Ericsson,
300 Holger Way
San Jose, CA
95134
Email: david.i.allan@ericsson.com

Nigel Bragg
Ciena Limited,
Ciena House
43-51 Worship Street
London  EC2A 2DX
Email: nbragg@ciena.com

Paul Unbehagen
Alcatel-Lucent
8742 Lucent Boulevard
Highlands Ranch, CO 80129, USA
Email: Paul.Unbehagen@alcatel-lucent.com

## 22. Intellectual Property Statement

## 23. Disclaimer of Liability

Network Working Group
INTERNET-DRAFT
Intended status: Proposed Standard

Donald Eastlake
Huawei
Ayan Banerjee
Dinesh Dutt
Cisco
Radia Perlman
Intel
Anoop Ghanwani
Brocade

Expires: August 1, 2011
February 2, 2011

# TRILL Use of IS-IS
<draft-ietf-isis-trill-05.txt>

## Abstract

The IETF has standardized the TRILL protocol, which provides transparent Layer 2 forwarding using encapsulation with a hop count and IS-IS link state routing. This document specifies the data formats and code points for the IS-IS extensions to support TRILL.

## Status of This Memo



Distribution of this document is unlimited. Comments should be sent to the ISIS and TRILL working group mailing lists: <isis-wg@ietf.org> and <rbridge@postel.org>.

Table of Contents

1. Introduction............................................3
1.1 Conventions used in this document......................3

2. TLV and sub-TLV Extensions to IS-IS for TRILL...........4
2.1 The Group Address TLV..................................4
2.1.1 The Group MAC Address sub-TLV........................4
2.2 Multi-Topology Aware Port Capability sub-TLVs..........6
2.2.1 The Special VLANs and Flags sub-TLV..................6
2.2.2 Enabled-VLANs sub-TLV................................7
2.2.3 Appointed Forwarders sub-TLV.........................8
2.3 Sub-TLVs for the Router Capability TLV.................9
2.3.1 The TRILL Version sub-TLV............................9
2.3.2 The Nickname sub-TLV................................10
2.3.3 The Trees sub-TLV...................................11
2.3.4 The Tree Identifiers Sub-TLV........................11
2.3.5 The Trees Used Identifiers Sub-TLV..................12
2.3.6 Interested VLANs and Spanning Tree Roots sub-TLV....12
2.3.7 The VLAN Group sub-TLV..............................15
2.4 MTU sub-TLV of the Extended Reachability TLV..........16
2.5 TRILL Neighbor TLV....................................16

3. The MTU PDUs...........................................19

4. Use of Existing PDUs and TLVs..........................20
4.1 TRILL IIH PDUs........................................20
4.2 Area Address..........................................20
4.3 Protocols Supported...................................20

5. Acknowledgements.......................................21

6. IANA Considerations....................................22
6.1 Allocations From Existing Registries..................22
6.2 New Sub-Registries Created............................23

7. Security Considerations................................25

8. References.............................................26
8.1 Normative References..................................26
8.2 Informative References................................27

Appendix A: Initial IS-IS PDU Registry....................28

# 1. Introduction

The IETF has standardized the TRILL protocol [RFCtrill], which provides transparent Layer 2 forwarding using encapsulation with a hop count and link state routing. TRILL provides optimal pair-wise forwarding without configuration, safe forwarding even during periods of temporary loops, and support for multipathing of both unicast and multicast traffic as well as supporting VLANs. Intermediate Systems (ISs) implementing TRILL can incrementally replace IEEE [802.1Q-2005] bridges.

This document, in conjunction with [RFCisisLayer2], specifies the data formats and code points for the IS-IS [ISO-10589] [RFC1195] extensions to support TRILL.

## 1.1 Conventions used in this document

The terminology and acronyms defined in [RFCtrill] are used herein with the same meaning.

Additional acronyms used in this document:

IIH - IS-IS Hello

IS - Intermediate System (for this document, all relevant intermediate systems are RBridges)

NLPID - Network Layer Protocol Identifier

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].

## 2. TLV and sub-TLV Extensions to IS-IS for TRILL

This section, in conjunction with [RFCisisLayer2], specifies the data formats and code points for the TLVs and sub-TLVs added to IS-IS to support the TRILL standard. Information as to the number of occurrences allowed, such as for a TLV in a PDU or set of PDUs or for a sub-TLV in a TLV, is provided in Section 6.

### 2.1 The Group Address TLV

The Group Address (GADDR) TLV, IS-IS TLV type 142 [TBD], is carried only in an LSP PDU and carries sub-TLVs that in turn advertise multicast group listeners. Section 2.1.1 below specifies a sub-TLV advertising listeners by MAC address. It is anticipated that additional sub-TLVS for additional address types such as IP addresses will be specified in other documents. The sub-TLVs under GADDR constitute a new series of sub-TLV types (see Section 6.2).

GADDR has the following format:

```
+-+-+-+-+-+-+-+-+
|Type=GADDR-TLV |                  (1 byte)
+-+-+-+-+-+-+-+-+
|   Length      |                  (1 byte)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|      sub-TLVs...
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

o Type: TLV Type, set to GADDR-TLV 142 [TBD].

o Length: variable depending on the sub-TLVs carried.

o sub-TLVs: The Group Address TLV value consists of sub-TLVs formatted as described in [RFC5305].

#### 2.1.1 The Group MAC Address sub-TLV

The Group MAC Address (GMAC-ADDR) sub-TLV is sub-TLV type number 1 within the GADDR TLV. In TRILL, it is used to advertise multicast listeners as specified in Section 4.5.5 of [RFCtrill]. It has the following format:

```
   +-+-+-+-+-+-+-+-+
   |Type=GMAC-ADDR |                  (1 byte)
   +-+-+-+-+-+-+-+-+
   |   Length      |                  (1 byte)
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |  RESV |     Topology-ID       |  (2 bytes)
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |  RESV |     VLAN-ID           |  (2 bytes)
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |Num Group Recs |                  (1 byte)
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                   GROUP RECORDS (1)                             |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                   .................                             |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                   GROUP RECORDS (N)                             |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

where each group record is of the form:

```
   +-+-+-+-+-+-+-+-+
   | Num of Sources|                  (1 byte)
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                   Group Address        (6 bytes)                |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                   Source 1 Address     (6 bytes)                |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                   Source 2 Address     (6 bytes)                |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                    .....                                        |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |                   Source M Address     (6 bytes)                |
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

- Type: GADDR sub-TLV Type, set to 1 (GMAC-ADDR).
- Length: Variable, minimum 5.
- RESV: Reserved. 4-bit fields that MUST be sent as zero and ignored on receipt.
- Topology-ID: This field is not used in TRILL, where it is sent as zero and ignored on receipt, but is included for use by other technologies.
- VLAN-ID: This carries the 12-bit VLAN identifier for all subsequent MAC addresses in this sub-TLV, or the value zero if no VLAN is specified.
- Number of Group Records: A 1-byte integer that is the number of

group records in this sub-TLV.

- o Group Record: Each group record carries the number of sources. It then has a 48-bit multicast address followed by 48-bit source MAC addresses. If the sources do not fit in a single sub-TLV, the same group address may be repeated with different source addresses in another sub-TLV of another instance of the Group Address TLV.

## 2.2 Multi-Topology Aware Port Capability sub-TLVs

TRILL makes use of the Multi-Topology aware Port Capability (MT-PORT-CAP) TLV as specified in [RFCisisLayer2]. The remainder of this section specifies the sub-TLVs that TRILL uses the MT-PORT-CAP TLV to transport.

### 2.2.1 The Special VLANs and Flags sub-TLV

In TRILL, a Special VLANs and Flags (VLAN-Flags) sub-TLV is carried in every IIH PDU. It has the following format:

```
+-+-+-+-+-+-+-+-+
|Type=VLAN Flags|                  (1 byte)
+-+-+-+-+-+-+-+-+
|   Length      |                  (1 byte)
+---------------+---------------+
|    Port ID                    |  (2 bytes)
+-------------------------------+
|     Sender Nickname           |  (2 bytes)
+--+--+--+--+-------------------+
|AF|AC|VM|BY|    Outer.VLAN     |  (2 bytes)
+--+--+--+--+-------------------+
|TR|R |R |R |    Desig.VLAN     |  (2 bytes)
+--+--+--+--+-------------------+
```

- o Type: sub-TLV Type, set to MT-PORT-CAP VLAN-Flags sub-TLV 1 [TBD].

- o Length: 8.

- o Port ID: An ID for the port on which the enclosing TRILL IIH PDU is being sent as specified in [RFCtrill] Section 4.4.2.

- o Sender nickname: If the sending IS is holding any nicknames as discussed in [RFCtrill] Section 3.7, one MUST be included here. Otherwise, the field is set to zero. This field is to support intelligent end stations that determine the egress IS (RBridge) for unicast data through a directory service or the like and need a nickname for their first hop to insert as the ingress nickname to correctly format a TRILL encapsulated data frame. See [RFCtrill] Section 4.6.2 point #8.

- Outer.VLAN: A copy of the 12-bit outer VLAN ID of the TRILL IIH frame containing this sub-TLV when that frame was sent, as specified in [RFCtrill] Section 4.4.5.

- Desig.VLAN: The 12-bit ID of the designated VLAN for the link as specified in [RFCtrill] Section 4.2.4.2.

- AF, AC, VM, BY, and TR: These flag bits have the following meaning each, if set to one, as specified in the listed section of [RFCtrill]:

| Bit | RFCtrill Section | Meaning if bit is one |
|---|---|---|
| AF | 4.4.2 | Originating IS believes it is Appointed Forwarder for the VLAN and port on which the containing IIH PDU was sent. |
| AC | 4.9.1 | Originating port configured as an access port (TRILL traffic disabled). |
| VM | 4.4.5 | VLAN Mapping detected on this link. |
| BY | 4.4.2 | Bypass pseudonode. |
| TR | 4.9.1 | Originating port configured as a trunk port (end station service disabled). |

- R: Reserved bit. MUST be sent as zero and ignored on receipt.

### 2.2.2 Enabled-VLANs sub-TLV

The optional Enabled-VLANs sub-TLV specifies the VLANs enabled for end station service at the port of the originating IS on which the Hello was sent as specified in [RFCtrill] Section 4.4.2.  It has the following format:

```
+-+-+-+-+-+-+-+-+
|Type=EnabledVLAN|                 (1 byte)
+-+-+-+-+-+-+-+-+
|   Length      |                  (1 byte)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| RESV  |  Start VLAN ID        |  (2 bytes)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| VLAN bit-map....
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

- o Type: sub-TLV Type, set to MT-PORT-CAP Enabled-VLANs sub-TLV 2 [TBD].

- o Length: Variable, minimum 3.

- o RESV: 4 reserved bits that MUST be sent as zero and ignored on receipt.

- o Start VLAN ID: The 12 bit VLAN ID that is represented by the high order bit of the first byte of the VLAN bit-map.

- o VLAN bit-map: The highest order bit indicates the VLAN equal to the start VLAN ID, the next highest bit indicates the VLAN equal to start VLAN ID + 1, continuing to the end of the VLAN bit-map field

If this sub-TLV occurs more than once in a Hello, the set of enabled VLANs is the union of the sets of VLANs indicated by each of the Enabled-VLAN sub-TLVs in the Hello.

### 2.2.3 Appointed Forwarders sub-TLV

The DRB on a link uses the Appointed Forwarders sub-TLV to inform other ISs on the link that they are the designated VLAN-x forwarder for one or more ranges of VLAN IDs as specified in Section 4.2.4 of [RFCtrill]. It has the following format:

```
+-+-+-+-+-+-+-+-+
|Type=ApptFwrdrs|                          (1 byte)
+-+-+-+-+-+-+-+-+
|   Length      |                          (1 byte)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   Appointment Information (1)         |  (6 bytes)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   .................                   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   Appointment Information (N)         |  (6 bytes)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

where each appointment is of the form:

```
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|       Appointee Nickname              |  (2 bytes)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| RESV  |        Start.VLAN             |  (2 bytes)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| RESV  |        End.VLAN               |  (2 bytes)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

- o Type: sub-TLV Type, set to MT-PORT-CAP Appointed Forwarders sub-TLV 3 [TBD].
- o Length: 6*n bytes where there are n appointments.
- o Appointee Nickname: The nickname of the IS being appointed a forwarder.
- o RESV: 4 bits that MUST be sent as zero and ignored on receipt.
- o Start.VLAN, End.VLAN: These fields are the VLAN IDs of the appointment range, inclusive. A VLAN's ID appears as both the start and end VLAN to specify that single VLAN. As specified in Section 4.4 of [RFCtrill], appointing an IS forwarder on a port for a VLAN not enabled on that port has no effect.

An IS's nickname may occur as appointed forwarder for multiple VLAN ranges by occurrences of this sub-TLV within the same or different MT Port Capability TLVs within an IIH PDU.

## 2.3 Sub-TLVs for the Router Capability TLV

The Router Capability TLV is specified in [RFC4971]. All of the sub-sections of this Section 2.3 below specify sub-TLVs that can be carried in the Router Capability TLV for TRILL.

### 2.3.1 The TRILL Version sub-TLV

The TRILL Version (TRILL-VER) sub-TLV indicates the maximum version of the TRILL standard supported. By implication, lower versions are also supported. If this sub-TLV is missing, the originating IS only supports the base version of the protocol [RFCtrill].

```
+-+-+-+-+-+-+-+-+
| Type          |                  (1 byte)
+-+-+-+-+-+-+-+-+
| Length        |                  (1 byte)
+-+-+-+-+-+-+-+-+
| Max-version   |                  (1 byte)
+-+-+-+-+-+-+-+-+
```

o  Type: Router Capability sub-TLV Type, set to 12 (TRILL-VER).

o  Length: 1.

o  Max-version: Set to maximum version supported.

### 2.3.2 The Nickname sub-TLV

The Nickname (NICKNAME) Router Capability sub-TLV carries information about the nicknames of the originating IS, along with information about its priority to hold those nicknames as specified in [RFCtrill] Section 3.7.3. Multiple instances of this sub-TLV may be carried.

```
+-+-+-+-+-+-+-+-+
|Type = NICKNAME|                         (1 byte)
+-+-+-+-+-+-+-+-+
|   Length      |                         (1 byte)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                NICKNAME RECORDS (1)                            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                   .................                            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                NICKNAME RECORDS (N)                            |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

where each nickname record is of the form:

```
+-+-+-+-+-+-+-+-+
| Nickname.Pri  |                  (1 byte)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     Tree Root Priority        |  (2 byte)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|           Nickname            |  (2 bytes)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

o  Type: Router Capability sub-TLV Type, set to 6 (NICKNAME).

o  Length: 5*N, where N is the number of nickname records present.

o  Nickname.Pri: An 8-bit unsigned integer priority to hold a nickname as specified in Section 3.7.3 of [RFCtrill].

o  Tree Root Priority: This is an unsigned 16-bit integer priority to be a tree root as specified in Section 4.5 of [RFCtrill].

o  Nickname: This is an unsigned 16-bit integer as specified in Section 3.7 of [RFCtrill].

### 2.3.3 The Trees sub-TLV

Each IS providing TRILL service uses the TREES sub-TLV to announce three numbers related to the computation of distribution trees as specified in Section 4.5 of [RFCtrill]. Its format is as follows:

```
+-+-+-+-+-+-+-+-+
|Type =  TREES  |                  (1 byte)
+-+-+-+-+-+-+-+-+
|  Length       |                  (1 byte)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Number of trees to compute    |  (2 byte)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Maximum trees able to compute |  (2 byte)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Number of trees to use        |  (2 byte)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

- Type: Router Capability sub-TLV Type, set to 7 (TREES).
- Length: 6.
- Number of trees to compute: An unsigned 16-bit integer as specified in Section 4.5 of [RFCtrill].
- Maximum trees able to compute: An unsigned 16-bit integer as specified in Section 4.5 of [RFCtrill].
- Number of trees to use: An unsigned 16-bit integer as specified in Section 4.5 of [RFCtrill].

### 2.3.4 The Tree Identifiers Sub-TLV

The tree identifiers (TREE-RT-IDs) sub-TLV is an ordered list of nicknames.  When originated by the IS that has the highest priority tree root, it lists the distribution trees that the other ISs are required to compute as specified in Section 4.5 of [RFCtrill].  If this information is spread across multiple sub-TLVs, the starting tree number is used to allow the ordered lists to be correctly concatenated. The sub-TLV format is as follows:

```
+-+-+-+-+-+-+-+-+
|Type=TREE-RT-IDs|                (1 byte)
+-+-+-+-+-+-+-+-+
|   Length      |               (1 byte)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|Starting Tree Number           |  (2 bytes)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   Nickname (K-th root)        |  (2 bytes)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   Nickname (K+1 - th root)    |  (2 bytes)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   Nickname (...)              |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

o  Type: Router Capability sub-TLV Type, set to 8 (TREE-RT-IDs).

o  Length: 2 + 2*n where n is the number of nicknames listed.

o Starting Tree Number: This identifies the starting tree number of the nicknames that are trees for the domain.  This is set to 1 for the sub-TLV containing the first list.  Other Tree-Identifiers sub-TLVs will have the number of the starting list they contain. In the event a tree identifier can be computed from two such sub-TLVs and they are different, then it is assumed that this is a transient condition that will get cleared.  During this transient time, such a tree SHOULD NOT be computed unless such computation is indicated by all relevant sub-TLVs present.

o  Nickname: The nickname at which a distribution tree is rooted.

### 2.3.5 The Trees Used Identifiers Sub-TLV

This Router Capability sub-TLV has the same structure as the Tree Identifiers sub-TLV specified in Section 2.3.4.  The only difference is that its sub-TLV type is set to 9 [TBD] (TREE-USE-IDs) and the trees listed are those that the originating IS wishes to use as specified in [RFCtrill] Section 4.5.

### 2.3.6 Interested VLANs and Spanning Tree Roots sub-TLV

The value of this Router Capability sub-TLV consists of a VLAN range and information in common to all of the VLANs in the range for the originating IS.  This information consists of flags, a variable length list of spanning tree root bridge IDs, and an appointed forwarder status lost counter, all as specified in the sections of [RFCtrill] listed with the respective information items below.

In the set of LSPs originated by an IS, the union of the VLAN ranges in all occurrences of this sub-TLV MUST be precisely the set of VLANs for which the originating IS is appointed forwarder on at least one port and the VLAN ranges in multiple VLANs sub-TLVs for an IS MUST NOT overlap unless the information provided about a VLAN is the same in every instance. However, as a transient state these conditions may be violated. If a VLAN is not listed in any INT-VLAN sub-TLV for an IS, that IS is assumed to be uninterested in receiving traffic for that VLAN. If a VLAN appears in more than one INT-VLAN sub-TLV for an IS with different information in the different instances, the following apply:

- If those sub-TLVs provide different nicknames it is unspecified which nickname takes precedence,
- The largest appointed forwarder status lost counter is used,
- The originating IS is assumed to be attached to a multicast IPv4 router for that VLAN if any of the INT-VLAN sub-TLVs assert that it is so connected and similarly for IPv6 multicast router attachment, and
- The root bridge lists from all of the instances of the VLAN for the originating IS are merged.

To minimize such occurrences, wherever possible, an implementation SHOULD advertise the update to a interested VLAN and spanning tree roots sub-TLV in the same LSP fragment as the advertisement that it replaces.  Where this is not possible, the two affected LSP fragments should be flooded as an atomic action. An IS that receives an update to an existing interested VLAN and spanning tree roots sub-TLV can minimize the potential disruption associated with the update by employing a hold-down timer prior to processing the update so as to allow for the receipt of multiple LSP fragments associated with the same update prior to beginning processing.

The sub-TLV layout is as follows:

```
   +-+-+-+-+-+-+-+-+
   |Type = INT-VLAN|                  (1 byte)
   +-+-+-+-+-+-+-+-+
   |   Length      |                  (1 byte)
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |   Nickname                    |  (2 bytes)
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+...+-+-+-+-+
   |   Interested VLANS                                      |  (4 bytes)
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+...+-+-+-+-+
   |   Appointed Forwarder Status Lost Counter               |  (4 bytes)
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+...+-+-+
   |         Root Bridges                                    |  (6*n bytes)
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+...+-+-+
```

o Type: Router Capability sub-TLV Type, set to 10 (INT-VLAN).

o Length: 10 + 6*n where n is the number of root bridge IDs.

o Nickname: As specified in [RFCtrill] Section 4.2.4.4, this field may be used to associate a nickname held by the originating IS with the VLAN range indicated. When not so used, it is set to zero.

o Interested VLANS: The Interested VLANs field is formatted as shown below.

```
     0    1    2    3     4 - 15      16 - 19     20 - 31
   +----+----+----+----+------------+----------+------------+
   | M4 | M6 |  R |  R | VLAN.start |   RESV   |  VLAN.end  |
   +----+----+----+----+------------+----------+------------+
```

   - M4, M6: These bits indicate, respectively, that there is an IPv4 or IPv6 multicast router on a link for which the originating IS is appointed forwarder for every VLAN in the indicated range as specified in [RFCtrill] Section 4.2.4.4 item 5.1.

   - R, RESV: These reserved bits MUST be sent as zero and are ignored on receipt.

   - VLAN.start and VLAN.end: This VLAN ID range is inclusive.  A range of one VLAN ID is indicated by setting them both to that VLAN ID value.

o Appointed Forwarder Status Lost Counter: This is a count of how many times a port that was appointed forwarder for the VLANs in the range given has lost the status of being an appointed forwarder as discussed in Section 4.8.3 of [RFCtrill]. It is initialized to zero at an IS when the zeroth LSP sequence number is initialized. No special action need be taken at rollover, the counter just wraps around.

- o Root Bridges: The list of zero or more spanning tree root bridge IDs is the set of root bridge IDs seen for all ports for which the IS is appointed forwarder for the VLANs in the specified range as discussed in [RFCtrill] Section 4.9.3.2. While, of course, only one spanning tree root could be seen on any particular port, there may be multiple ports in the same VLAN connected to differed bridged LANs with different spanning tree roots.

An INT-VLAN sub-TLV asserts that the information provided (multicast router attachment, appointed forwarder status lost counter, and root bridges), is the same for all VLANs in the range specified. If this is not the case, the range MUST be split into subranges meeting this criteria. It is always safe to use sub-TLVs with a "range" of one VLAN ID but this may be too verbose.

### 2.3.7 The VLAN Group sub-TLV

The VLAN Group Router Capability sub-TLV consists of two or more VLAN IDs as specified in [RFCtrill] Section 4.8.4. This sub-TLV indicates that shared VLAN learning is occurring at the announcing IS between the listed VLANs. It is structured as follows:

```
+-+-+-+-+-+-+-+-+
|Type=VLAN-GROUP|                  (1 byte)
+-+-+-+-+-+-+-+-+
|   Length      |                  (1 byte)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| RESV  |  Primary VLAN ID      |  (2 bytes)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| RESV  |  Secondary VLAN ID    |  (2 bytes)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  more Secondary VLAN IDs ...     (2 bytes each)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

- o Type: Router Capability sub-TLV Type, set to 13 (VLAN-GROUP).
- o Length: 4 + 2*n, where n is the number of secondary VLAN ID fields, which may be zero.
- o RESV: a 4-bit field that MUST be sent as zero and ignored on receipt.
- o Primary VLAN-ID: This identifies the primary VLAN-ID.
- o Secondary VLAN-ID: This identifies a secondary VLAN in the VLAN Group.

- o more Secondary VLAN IDs: zero or more bytes pairs with the top four bits an RESV field and the low 12 bits a VLAN-ID.

## 2.4 MTU sub-TLV of the Extended Reachability TLV

The MTU sub-TLV is used to optionally announce the MTU of a link as specified in [RFCtrill] Section 4.2.4.4. It occurs within the Extended Reachability TLV (type #22).

```
+-+-+-+-+-+-+-+-+
| Type = MTU    |                  (1 byte)
+-+-+-+-+-+-+-+-+
|   Length      |                  (1 byte)
+-+-+-+-+-+-+-+-+
|F|  Reserved   |                  (1 byte)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|              MTU              |  (2 bytes)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

- o Type: Extended Reachability sub-TLV Type, set to MTU sub-TLV 28 [TBD].
- o Length: 3.
- o F: Failed. This bit is a one if MTU testing failed on this link at the required campus-wide MTU.
- o Reserved: 7 bits that MUST be sent as zero and ignored on receipt.
- o MTU: This field is set to the largest successfully tested MTU size for this link, or zero if it has not been tested, as specified in Section 4.3.2 of [RFCtrill].

## 2.5 TRILL Neighbor TLV

The TRILL Neighbor TLV is used in TRILL IIH PDUs (see Section 4.1 below) in place of the IS Neighbor TLV, as specified in Section 4.4.2.1 of [RFCtrill] and [RFCadj]. The structure of the TRILL Neighbor TLV is as follows:

```
+-+-+-+-+-+-+-+-+
|Type=TRILLNeigh|                  (1 byte)
+-+-+-+-+-+-+-+-+
|   Length      |                  (1 byte)
+-+-+-+-+-+-+-+-+
|S|L|  RESV     |                  (1 byte)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                Neighbor RECORDS (1)                             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                   .................                             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                Neighbor RECORDS (N)                             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

The information present for each neighbor is as follows:

```
+-+-+-+-+-+-+-+-+
|F|  RESV       |                (1 bytes)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|       MTU                     |  (2 bytes)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+...+-+-+-+-+-+-+-+
|   MAC Address                                         | (6 bytes)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+...+-+-+-+-+-+-+-+
```

o  Type: TLV Type, set to TRILL Neighbor TLV 145 [TBD].

o  Length: 1 + 9*n, where n is the number of neighbor records which may be zero.

o  S: Smallest flag.  If this bit is a one, then the list of neighbors includes the neighbor with the smallest MAC address considered as an unsigned integer.

o  L: Largest flag.  If this bit is a one, then the list of neighbors includes the neighbor with the largest MAC address considered as an unsigned integer.

o  RESV: These seven bits are reserved use and MUST be sent as zero and ignored on receipt.

o  F: failed.  This bit is a one if MTU testing to this neighbor failed at the required campus-wide MTU (see [RFCtrill] Section 4.3.1).

o  MTU: This field is set to the largest successfully tested MTU size for this neighbor or zero if it has not been tested.

o  MAC Address: The MAC address of the neighbor as in the IS Neighbor TLV (#6).

As specified in [RFCtrill] Section 4.4.2.1 and [RFCadj], all MAC address may fit into one TLV, in which case both the S and L flags would be set to one in that TLV. Otherwise, the highest MAC address in a TRILL Neighbor TLV with the L flag zero MUST also appear as a MAC address in some other TRILL Neighbor TLV (possibly in a different TRILL IIH PDU). And the lowest MAC address in a TRILL Neighbor TLV with the S flag zero MUST also appear in some other TRILL Neighbor TLV (possibly in a different TRILL IIH PDU). If an RBridge believes it has no neighbors, it MUST send a TRILL Neighbor TLV with an empty list of neighbor RECORDSs, which will have both the S and L bits on.

## 3. The MTU PDUs

Two PDUs are added to IS-IS, the MTU-probe and MTU-ack PDUs. They are used to optionally determine the MTU on a link between ISs as specified in [RFCtrill] Section 4.3.2.

The MTU PDUs have the IS-IS PDU common header (up through the Maximum Area Addresses byte) with two new PDU Type numbers, one each, as listed in Section 6. They also have a 20-byte common fixed MTU PDU header as shown below.

```
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|    PDU Length                 |  (2 bytes)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+.....+-+-+
|    Probe ID                               (6 bytes)               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+.....+-+-+
|    Probe Source ID                        (6 bytes)               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+.....+-+-+
|    Ack Source ID                          (6 bytes)               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+.....+-+-+
```

As with other IS-IS PDUs, the PDU length gives the length of the entire IS-IS packet starting with and including the IS-IS common header.

The Probe ID field is an arbitrary 48-bit quantity set by the IS issuing an MTU-probe and copied by the responding IS into the corresponding MTU-ack. For example, an IS creating an MTU-probe could compose this quantity from a port identifier and probe sequence number relative to that port.

The Probe Source ID is set by an IS issuing an MTU-probe to its System ID and copied by the responding IS into the corresponding MTU-ack.

The Ack Source ID is set to zero in MTU-probe PDUs. An IS issuing an MTU-ack sets this field to its System ID.

The TLV area follows the MTU PDU header area. This area MAY contain an Authentication TLV and MUST be padded to the exact size being tested with the Padding TLV. Since the minimum size of the Padding TLV is 2 bytes, it would be impossible to pad to exact size if the total length of the required information bearing fixed fields and TLVs added up to 1 byte less than the desired length; however, the length of the fixed fields and substantive TLVs for MTU PDUs will be quite small compared with their minimum length (minimum 1470 byte MTU on an 802.3 link for example), so this will not be a problem.

## 4. Use of Existing PDUs and TLVs

The sub-sections below provide details of TRILL use of existing PDUs and TLVs.

### 4.1 TRILL IIH PDUs

The TRILL IIH PDU is the variation of the LAN IIH PDU used by the TRILL protocol. Section 4.4 of the TRILL standard [RFCtrill] specifies the contents of the TRILL IIH and how its use differs in TRILL from Layer 3 LAN IIH PDU use. The adjacency state machinery for TRILL neighbors is specified in Section 4.4 of [RFCtrill] and in [RFCadj].

In a TRILL IIH PDU the IS-IS Common Header and the fixed PDU Header are the same as a Level 1 LAN IIH PDU. The Maximum Area Addresses octet in the Common Header MUST be set to 0x01.

The IS-IS Neighbor TLV (#6) is not used in a TRILL IIH and is ignored if it appears there.  Instead, TRILL IIH PDUs use the TRILL Neighbor TLV (see Section 2.6).

### 4.2 Area Address

TRILL uses a fixed zero Area Address as specified in [RFCtrill] Section 4.2.3. This is encoded in a four byte Area Address TLV (TLV #1) as follows:

```
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   0x01, Area Address Type     |   (1 byte)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   0x02, Length of Value       |   (1 byte)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   0x01, Length of Address     |   (1 byte)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   0x00, zero Area Address     |   (1 byte)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

### 4.3 Protocols Supported

NLPID 0xC0 has been assigned to TRILL [RFCnlpid]. A Protocols Supported TLV (#129, [RFC1195]) including that value MUST appear in TRILL IIH PDUs and LSP number zero PDUs.

## 5. Acknowledgements

The authors gratefully acknowledge the contributions and review by the following: Mike Shand, Stewart Bryant, Dino Farinacci, Les Ginsberg, Sam Hartman, Dan Romascanu, Dave Ward, and Russ White, particularly the detailed and helpful comments by Mike Shand.

## 6. IANA Considerations

IANA is requested to allocate the existing registry code points listed in Section 6.1 and create two new registries with their initial contents as described in Section 6.2.

RFC Editor Note: In a number of cases in this draft, a specific code point value appears followed by " [TBD]". This indicates that there is a very small chance that the IANA / Expert process might change the code point. The editing action of removing the " [TBD]" is expected to be appropriate. Where a "[TBD]" appears without a sepcifc code point value, it should be replaced by the allocated value. Please remove this note prior to publication.

### 6.1 Allocations From Existing Registries

This document specifies two new IS-IS TLV types, namely the Group Address TLV (GADDR, type 142) and the TRILL Neighbor TLV (type 145). The PDUs in which these TLVs are permitted for TRILL are shown in the table below along with the section of this document where they are discussed. The final "NUMBER" column indicates the permitted number of occurrences of the TLV in their PDU, or set of PDUs in the case of LSP, which in these two cases is "*" indicating that the TLV MAY occur 0, 1, or more times.

It is requested that these two code points be added to the IANA IS-IS TLV registry (ignoring the "Section" and "NUMBER" columns that are irrelevant to that registry).

| | Section | TLV# | IIH | LSP | SNP | NUMBER |
|---|---|---|---|---|---|---|
| GADDR-TLV | 2.1 | 142 | - | X | - | * |
| TRILL Neighbor TLV | 2.5 | 145 | X | - | - | * |

This document specifies eleven new sub-TLVs from existing sub-TLV sequences, namely VLAN-FLAGS, Enabled-VLANs, AppointedFwrdrs, TRILL Version (TRILL-VER), Nickname, TREES, TREE-RT-IDs, TREE-USE-IDs, INT-VLAN, VLAN-GROUP, and MTU. The TLVs in which these sub-TLVs occur are shown in the table below along with the Section of this document where they are discussed.

Those sub-TLVs with an "X" in the column labeled "MT Port Capabil." are sub-TLVs of TLV #TBD [RFCisisLayer2], the MT-PORT-CAP TLV. Those sub-TLVs with an "X" in the column labeled "Router Capabil." are sub-TLVs of TLV #242, IS-IS Router Capabilities TLV. Those sub-TLVs with an "X" in the column labeled "Extended IS Reach" are sub-TLVs of TLV #22, the Extended IS Reachability TLV.

The final "NUM" column indicates the permitted number of occurrences of the sub-TLV cumulatively within all occurrences of their TLV in that TLV’s carrying PDU (or set of PDUs in the case of LSP), as follows:

- 0-1 = MAY occur zero or one times. If it occurs more than once, results are unspecified.
- 1 = MUST occur exactly once. If absent, the PDU is ignored. If it occurs more than once, results are unspecified.
- * = MAY occur 0, 1, or more times.

The values in the "Section" and "NUM" columns are irrelevant to the IANA sub-registries.

| | Section | sub-TLV# | MT Port Capabil. | Router Capabil. | Extended IS Reach | NUM |
|---|---|---|---|---|---|---|
| VLAN-FLAGS | 2.2.1 | 1 | X | - | - | 1 |
| Enabled-VLANs | 2.2.2 | 2 | X | - | - | * |
| AppointedFwrdrs | 2.2.3 | 3 | X | - | - | * |
| NICKNAME | 2.3.2 | 6 | - | X | - | * |
| TREES | 2.3.3 | 7 | - | X | - | 0-1 |
| TREE-RT-IDs | 2.3.4 | 8 | - | X | - | * |
| TREE-USE-IDs | 2.3.5 | 9 | - | X | - | * |
| INT-VLAN | 2.3.6 | 10 | - | X | - | * |
| TRILL-VER | 2.3.1 | 12 | - | X | - | 0-1 |
| VLAN-GROUP | 2.3.7 | 13 | - | X | - | * |
| MTU | 2.4 | 28 | - | - | X | 0-1 |

## 6.2 New Sub-Registries Created

This document creates two new IS-IS PDUs, namely the MTU-PROBE-PDU, and MTU-ACK-PDU, as described in Section 3. IANA will assign new PDU types to these PDUs and reflect them in a newly created PDU registry (see Appendix A). [suggested PDU values below]

```
MTU-PROBE-PDU     PDU Number: TBD (23 suggested)
MTU-ACK-PDU       PDU Number: TBD (28 suggested)
```

IANA is requested to create a new sub-TLV IS-IS sub-registry for sub-TLVs within the Group Address (GADDR) TLV and specifies an initial sub-TLV within that registry, namely Group MAC Address (GMAC-ADDR), sub-TLV #1. The GMAC-ADDR sub-TLV may occur 0, 1, or more times in a GADDR TLV.

The initial sub-registry is shown below.

Registry Name: IS-IS Group Address Type Codes for TLV 10
Reference: This document

Registration Procedures: Expert Review [RFC5226]

Registry:

| Value | Group Address Type Code | Reference |
| --- | --- | --- |
| 0 | Reserved | This document |
| 1 | MAC Address | This document |
| 2-254 | Unassigned | This document |
| 255 | Reserved | This document |

## 7. Security Considerations

For general TRILL protocol security considerations, see the TRILL base protocol standard [RFCtrill].

This document raises no new security issues for IS-IS. IS-IS security may be used to secure the IS-IS messages discussed here. See [RFC5304] and [RFC5310]. Even when the IS-IS authentication is used, replays of Hello packets can create denial-of-service conditions; see [RFC6039] for details. These issues are similar in scope to those discussed in Section 6.2 of [RFCtrill], and the same mitigations may apply.

# 8. References

Normative and informative references for this document are given below.

## 8.1 Normative References

[ISO-10589] - ISO/IEC 10589:2002, Second Edition, "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)", 2002.

[RFC1195] - Callon, R., "Use of OSI IS-IS for Routing in TCP/IP and Dual Environments", 1990.

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

[RFC4971] - Vasseur, JP. and N. Shen, "Intermediate System to Intermediate System (IS-IS) Extensions for Advertising Router Information", 2007.

[RFC5226] - Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 5226, May 2008.

[RFC5305] - Li, T. and H. Smit, "IS-IS Extensions for Traffic Engineering", 2008.

[RFCadj] - "RBridges: Adjacency", draft-ietf-trill-adj, work in progress.

[RFCisisLayer2] - Banerjee, A., and D. Ward, "Extensions to IS-IS for Layer-2 Systems", draft-ietf-isis-layer2-09.txt, work in progress.

[RFCnlpid] - Eastlake, D., "IANA Considerations for Network Layer Protocol Identifiers", draft-eastlake-nlpid-iana-considerations-04.txt, in RFC Editor’s queue.

[RFCtrill] - Perlman, R., D. Eastlake, D. Dutt, S. Gai, and A. Ghanwani, "RBridges: Base Protocol Specification", draft-ietf-trill-rbridge-protocol-16.txt, in RFC Editor’s queue.

## 8.2 Informative References

[802.1Q-2005] "IEEE Standard for Local and metropolitan area networks / Virtual Bridged Local Area Networks", 802.1Q-2005, 19 May 2006.

[RFC5304] - Li, T. and R. Atkinson, "IS-IS Cryptographic Authentication", RFC 5304, October 2008.

[RFC5310] - Bhatia, M., Manral, V., Li, T., Atkinson, R., White, R., and M. Fanto, "IS-IS Generic Cryptographic Authentication", RFC 5310, February 2009.

[RFC6039] - Manral, V., Bhatia, M., Jaeggli, J., and R. White, "Issues with Existing Cryptographic Protection Methods for Routing Protocols", RFC 6039, October 2010.

## Appendix A: Initial IS-IS PDU Registry

The following is the suggested initial IS-IS PDU Registry before MTU-PROBE-PDU and MTU-ACK-PDU, which should be added with this document as REFERENCE:

Registry Name: IS-IS PDUs
Reference: This document
Registration Procedures: IETF Review [RFC5226]

| MNEMONIC | PDU# | REFERENCE |
|---|---|---|
| Unassigned | 0-14 | |
| L1-LAN-HELLO-PDU | 15 | [ISO-10589] |
| L2-LAN-HELLO-PDU | 16 | [ISO-10589] |
| P2P-HELLO-PDU | 17 | [ISO-10589] |
| L1-LSP-PDU | 18 | [ISO-10589] |
| Unassigned | 19 | |
| L2-LSP-PDU | 20 | [ISO-10589] |
| Unassigned | 21-23 | |
| L1-CSNP-PDU | 24 | [ISO-10589] |
| L2-CSNP-PDU | 25 | [ISO-10589] |
| L1-PSNP-PDU | 26 | [ISO-10589] |
| L2-PSNP-PDU | 27 | [ISO-10589] |
| Unassigned | 28-31 | |

Authors' Addresses

Donald Eastlake
Huawei
155 Beaver Street
Milford, MA 01757 USA

Phone: +1-508-333-2270
email: d3e3e3@gmail.com

Ayan Banerjee
Cisco Systems
170 West Tasman Drive
San Jose, CA 95134 USA

Email: ayabaner@cisco.com

Dinesh Dutt
Cisco Systems
170 West Tasman Drive
San Jose, CA 95134-1706 USA

Phone: +1-408-527-0955
Email: ddutt@cisco.com

Radia Perlman
Intel Labs
2200 Mission College Blvd.
Santa Clara, CA 95054-1549 USA

Phone: +1-408-765-8080
Email: Radia@alum.mit.edu

Anoop Ghanwani
Brocade Communications Systems
1745 Technology Drive
San Jose, CA 95110 USA

Phone: +1-408-333-7149
Email: anoop@brocade.com

Copyright and IPR Provisions