Provider Edge Bridge Spanning Tree

Mick Seaman

This note is a precursor to suggesting text for .1ad to support customer spanning trees across a provider bridged network. It describes the handling of RSTP BPDUs in the CVLAN-aware component of a Provider Edge Bridge†1.

1. Connecting LANs or VLANs?

Before deciding how RSTP should operate in the CVLAN-aware components of Provider Edge Bridges†2, it is important to be clear what sort of connectivity is being supported. Is the customer's spanning tree being supported to avoid potential loops:

a) amongst any and all VLANs, together with traffic not assigned to any VLAN?

or

b) amongst each of a number of VLANs?

In the latter case loops that involve taking traffic from one VLAN and putting it on another are outside the scope of the spanning tree operation. They may be subject to their own loop prevention or mitigation mechanisms, as is routing with its hop count. Consistency checks that guard against 'cross-wiring' VLANs may be implemented.

Supporting case (a) amongst provider edge bridges is not likely to win customer acclaim. It would mean that the connectivity between customer sites would be reduced to a spanning tree of service instances within the provider network. This is hardly ever the customer intent, and might be quite surprising, since separate customer VLAN identifiers are being used to identify the service instances. However (a) is worth mentioning since there will inevitably be some customers who are not running VLANs within their sites, and who attempt to setup connectivity that is meant to use redundant (redundant to the customer that is) service instances within a provider's network but are accessed through the same customer edge port on a provider edge bridge†3.

Case (b) is the interesting one. If each customer VLAN can only be mapped to and from a single service instance†4, the operation of RSTP (as described below) between the provider edge bridges will not block communication that use any of the service instances, unless there is an alternate path to that provided by the service provider. The customer may be operating RSTP, with SST Bridges supporting many VLANs over that single tree, or MSTP.

To provide the desired result, RSTP is enhanced to support VLAN-sensitive operation. This is a subset of the techniques originally proposed in 'Spanning Vines'.

†2See Section A and Figure 6 for a summary of Provider Bridge Architecture.

†3OK, it may not make a lot of sense since the interface to the provider edge bridge is not redundantly connected, but someone is going to do this. We had better not try to fix the 'bug' reported.

†4As required to avoid the potential of 'U-turns' within the CVLAN,

2. VLAN-sensitive RSTP operation

The proposed VLAN-sensitive mode of operation is an enhancement to RSTP that allows connectivity for any given VLAN through a Bridge Port that the spanning tree algorithm has calculated to be an Alternate Port, instead of the Root Port. The selected Port (whether the calculated Root Port or a substituted Alternate Port) is referred to as the Rootward Port for that VLAN†5 †6. The algorithm used to select it depends on the use scenario, and is a local decision (i.e. other bridges do not need to know it, and can make their own independent choice†7). The RSTP protocol, at least in terms of the contents of BPDUs sent, and when they are sent†8 is unaffected†9. In the case of edge bridge connectivity across the provider network, the highest priority port that actually supports connectivity for the VLAN is chosen.

This algorithm works well for provider edge bridges with and without the 'no U turn' restriction, i.e. only one permitted service instance per C-VLAN.

Each C-VLAN bridge component treats the Customer Edge Ports and all the internal Provider Edge Ports (one per provided service instance, see Figure 6) as ordinary ports in the spanning tree algorithm. All the VLANs that the C-bridge will forward can pass through the Customer Edge Port, so VLAN-sensitivity has no effect when that is chosen as the Root Port. When one of the internal ports is chosen as Root Port, the highest priority internal port capable of forwarding traffic for each C-VLAN is chosen as its Rootward Port.

†5The 'Leafward' Ports for a given VLAN are simply those Designated Ports (as selected by RSTP) that are allowed to forward traffic for that VLAN. In the case of a general mesh VLAN-sensitive RSTP operation provides less functionality than MSTP (apart from lacking the complete flexibility of independent trees, it cannot signal topology changes for 'off tree' VLANs and needs supplementing with MVRP), but has much simpler configuration requirements. In a wide variety of scenarios it can be autoconfigured through application of a general Rootward Port selection algorithm, with an optional policy component. In the simple case of Provider Edge Bridges communicating over a Provider Bridge Network, all the required functionality is provided.

†6Suggestions for a better name than 'Rootward' would be welcome. I started by using the name 'Up Port'. However, while the Rootward Port is 'Up' in the sense that computer science trees are normally drawn, i.e. with their roots uppermost, it is 'down' in the way that most campus networks are described i.e. the port that leads down not up the riser. I tried 'Uppermost', but that doesn't work for a number of reasons.

†7Including simply choosing the Root Port, whether the reception of a given VLAN on the port is possible or not.

†8To within a few milliseconds.

†9However the additional functionality provided is essential in the Provider Edge Bridge application, and would seem to warrant a new acronym at least so VLAN-sensitive operation capable operation can easily be distinguished from plain RSTP. VSTP is suggested.

†1Spanning tree handling to provide loop free connectivity within the provider network supports, but is separate from, loop free connectivity for customer network that can use a provider network for part of its connectivity. The provider network is likely to base its internal connectivity on MSTP. Where an S-tagged interface is provided to a customer, it may be useful to send the customer the provider's MSTP BPDUs, as that allows redundantly connected customer equipment to (a) provide continuous service in the face of temporary partitions within the edge of the providers network, rather than just protecting the physical link from the provider (b) select a best link to the provider for any given S-VLAN. However the customer never participates in the provider's own spanning tree, and the provider sets the restrictedRole and restrictedTcn flags for the interface port to make sure the customer cannot affect or disrupt that tree. Similarly the provider sets those parameters for S-VLAN component ports that attach to the C-VLAN components in a Provider Edge Bridge as a defensive measure.

The fine details include:

a) The dynamics of moving connectivity for any given VLAN from one Rootward Port to another.

   Forwarding for a VLAN through a new Rootward Port cannot be permitted until the prior Rootward Port is blocked.

b) Handling Topology Change Notifications.

They are more easily specified if each Provider Edge Bridge locally enforces mapping of each C-VLAN to at most one service instance The rest of this note assumes that restriction†1, with the following consequences:

1) Data is never forwarded between the internal ports seen by the CVLAN component (though the absence of such forwarding is not a matter for spanning tree port states, but for the filtering provided by VLAN Registration Entries).
2) The VLAN-sensitive RSTP specification can be phrased purely in terms of ports, without reference to individual VLANs.
3) There is no requirement to propagate topology change notifications from one internal port to another, thus avoiding any difficulty arising from the absence of a way to associate a topology change with particular VLANs, and thus prevent it looping amongst internal ports.

3. Configuration parameters

To minimize the chance of interfering with the customer's spanning tree configuration, as opposed to supporting its operation, the following spanning tree parameter settings are recommended:

a) The Bridge Identifier Priority/Bridge Priority (802.1D-2004 Clause 17.14 and Table 17-2) should be set to 61,440. This sets the priority part of the Bridge Identifier (the most significant 4 bits, see 802.1D Clause 9.2.5) to hex F.

   The following 12 bits (the Bridge Identifier system ID extension) should be set to hex FFF.

   These settings minimize the chance of the C-VLAN component of a Provider Edge Bridge becoming the Root Bridge for the customer's spanning tree.

b) The Port Priority (802.1D-2004 Clause 17.14 and Table 17-2) should be set to 32. This sets the priority part of the Port Identifier (the most significant 4 bits) to hex 2, a higher priority than the default (128, or hex 8).

c) The Port Path Cost values for Customer Edge Ports are set to their recommended values (802.1D-2004 Table 17-3) based on the transmission speeds of the Customer Edge Port and of the service instances corresponding to each of the Provider Edge Ports.

d) The Port Path Cost values for Provider Edge Ports are set to 128. This corresponds to the 802.1D recommendations for a speed above 10Gb/s. This value is chosen to minimize the chance of an inconvenient partition in the spanning tree topology (see Figure 4). BPDUs passing through the provider network will have a cost added for both the Customer Edge and Provider Edge Ports, so will not contain unusual values.

†1I should not overstate the difficulty of describing the general case. If it is wanted it can be done with a modest addition to this note.

4. Examples

The following examples show three Customer Edge Ports, and their associated C-VLAN components (A, B, C), attached to and interconnecting a customer's LANs (La, Lb, Lc) with service instances 1, 2, 3, 4 that correspond to disjoint sets of VLANs (also referred to as 1, 2, 3, and 4†2).

A, B, and C are using VLAN-sensitive RSTP. The customer's bridges use standard RSTP. Port Roles are shown using the conventional notation†3, with Rootward Ports shown as Alternate Ports that are not Discarding frames.

Assuming that the customer's Root Bridge is attached to La, the Port Roles, Port States, and connectivity will be as shown in Figure 1.

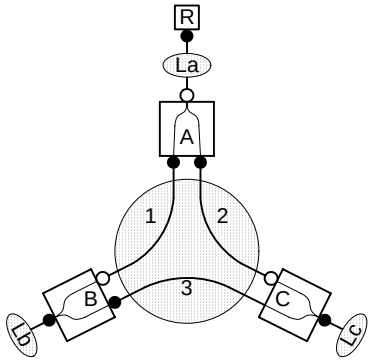

Figure 1—Pt-to-pt connections

Frames are forwarded from La to Lb using service instance 1, from La to Lc using 2, and from Lb to Lc using 3. Frames received at B on 1 are not forwarded to C using 3 because the C-VLAN sets for 1 and 3 are disjoint.

Of course it is entirely possible that the customer has attached routers and no bridges, and is therefore not running RSTP at all. In this case the configuration would resolve as in Figure 2, providing full connectivity, but still being capable of protecting against a loop caused by accidental addition of bridging between the customer LANs.

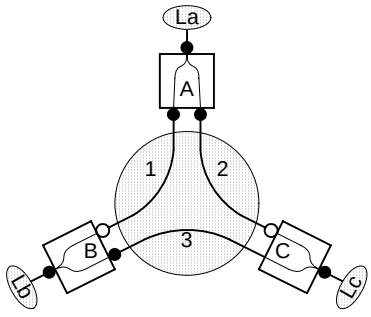

Figure 2—Operation with customer spanning tree

†2It can be the cases that one (and no more) VLANs per service instance have different C-VIDs at different C-VLAN bridge components. This detail has no material effect on the discussion and has been omitted.

†3802.1D-2004 Figure 17-1.

Figure 3 shows such bridge connectivity, added to the network of Figure 1.

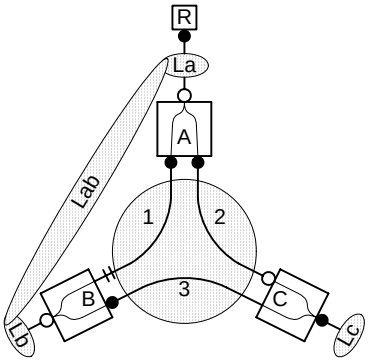

Figure 3—Loop prevention

The customer's LAN between A and B is assumed to provide connectivity for all C-VLANs (no better information being available). Assuming this to be true, there is no loss in connectivity. The C-VLANs for service instance 1 are carried by Lab. Service instance 2 still provides connectivity for its C-VLANs from A to C, and 3 from B to C. There is a chance that the Customer Edge Port for B would break the loop, rather than the Provider Edge Port for service instance 1, as Figure 4 shows.

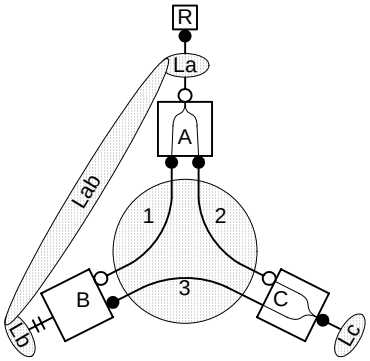

Figure 4—Sub-optimal loop prevention

The Provider Edge Port Path Cost and Customer Edge Port Priority are chosen to minimize the chance of this occurring, and to make it easier for the customer to choose between blocking service instance 1 connectivity and blocking connectivity somewhere in Lab.

Of course multiple service instances can be provided between two customer sites, conveying disjoint sets of C-VLANs. One useful combination is to use a multipoint service instance to meet bridged connectivity requirements†1, with point-to-point services for routers. Figure 5 provides an example.

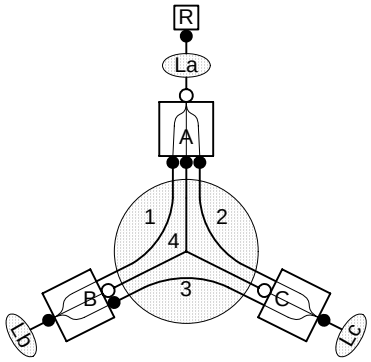

Figure 5—Services for bridging and routing

5. Detailed specification

The detailed specification†2 uses the Provider Bridge Architecture summarized in Section A and is based on the RSTP state machines of 802.1D-2004 (as updated by the maintenance corrections that will appear in the P802.1Q-REV/D2.0). The following procedures and state machine conditions are changed:

1) The updtRolesTree() procedure (17.21.25) is modified to assign the Port Role of Rootward Port to each of the Provider Edge Ports iff:
   a) one of the Provider Edge Ports has been selected as the Root Port; and
   a) the Port would otherwise be assigned an Alternate Port Role.
2) The global transition to the Port Role Transitions machine ROOT PORT state is extended to include a selectedRole of RootwardPort.
3) The procedures setReRootTree(), setSyncTree(), and setTcPropTree() each currently set a variable (reRoot, sync, and tcProp respectively) true for all (all other in the case of tcProp) ports of the Bridge. Each of these procedures is modified so that if the Root or Rootward Port in questions is:
   a) the Customer Edge Port, then the variable is set true for all (all other ports); but if it is
   b) a Provider Edge Port (i.e. a per service instance port internal to the Provider Edge Bridge), then the variable is set true for the Customer Edge Port (and for the originating port if 'all' rather than 'all other' was originally specified.
4) The state machine condition allSynced currently requires synced to be true for all ports other than the Root Port. The definition of the condition is changed so that there is an independent value of allSynced for each Port of the Bridge (the condition is only used in the Port Role Transitions state machine (PRT) in the Root/Rootward and Alternate states), and if the Port is:
   a) the Customer Edge Port, then synced is true for all other ports;
   b) a Provider Edge Port, then synced is true for the Customer Edge Port.

†1This can support GVRP/MVRP to dynamically adjust the extent of connectivity for the bridged C-VLANs.

†2The following is not the detailed specification, merely a description of what it needs to contain.

5) Similarly, the definition of the state machine condition reRooted is changed so that there is an independent value for each Port of the Bridge and if the Port is:

   a) the Customer Edge Port, then rrWhile is zero for all other ports;

   b) a Provider Edge Port, then rrWhile is zero for the Customer Edge Port.

NOTE—The changes (3), (4), and (5) above are those that would be required if the MSTP (rather than the RSTP) specification were to be used, with a separate Tree being associated with each Provider Edge Port, and the only port supporting more than one Tree being the Provider Edge Port. However identifying each Rootward Port and the Root Port of the relevant tree doesn't quite work in the rest of the specification. In particular Role Selection (updtRoleTree()) needs to know the difference between a Root Port and a Rootward Port. There is also a curiosity in describing the connectivity for separate C-VLANs as separate Trees, as some have neither a Root or a Rootward Port, but comprise only connected Designated Ports. This is because the only choice for Rootward Port is the Root Port, and that it a Provider Edge Port and does not carry traffic for other C-VLANs—bridging between C-VLANs using the C-VLAN component is not permitted as a substitute for acquiring sufficient service instance connectivity.

6. Additional considerations

Spanning tree protocols can only resolve, rather than prevent, loops that arise as a result of physically connecting two LANs that already appear to be operational. It is fortunate that manageable repeaters are not extensively used today. However a similar situation can arise when one spanning tree is running over another, particularly if the lower tree is slower to converge and provide connectivity than the upper. It is possible that a Designated Port for the upper tree concludes that it can start forwarding frames just before the lower tree connects two such ports. This would result in a temporary loop.

The potential for this effect is diminished by use of the Proposal/Agreement mechanism in RSTP and MSTP. A real lower tree is likely to be running over point-to-point links and will use the fast transition. Upper tree Designated Ports that transition to forwarding because they are the only ports attached to a service instance, or are a subset of the ports potentially attached to multipoint service instance, will use the forward delay timer to delay the transition to forwarding.

However we should still consider Norm Finn's proposal to extend the notion of Edge Port to include an explicit "Not Edge" Port as well as a don't know condition. Call such a port an 'Interior Port' for the time being. An 'Interior Port' would have the characteristic that the transition to forwarding would be inhibited if BPDUs were not seen from another Bridge. This would prevent loops arising from a long 'break' followed by a 'make' at the S-VLAN/service instance level†1. A C-VLAN component bridge is not necessarily in communication with another C-VLAN component, since the other service interface may be port-based or S-tagged, however the 'Interior Port' setting could be easily assigned at service provisioning time.

†[1]The service instance could be supported by MPLS or some other L2VPN.

A. Provider Edge Bridge Architecture

Customer equipment connected to a Provider Edge Bridge selects between S-VLANs by C-VLAN tagging transmitted frames. The operation of the Provider Edge Bridge is modeled as comprising two component bridges, as illustrated in Figure 6.

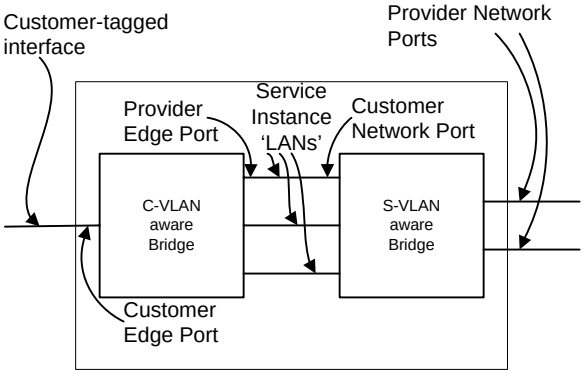

Figure 6—Provider Edge Bridge

The two bridges are connected by an number of internal LANs, one per S-VLAN. These are treated just like real LANs by each of the bridges, with a Bridge Port attaching to each.