PBT/D1.0
May 10, 2006

General Discussion of Provider Backbone Transport in 802.1ah Networks

Paul Bottorff
Tel: 408-495-3365 Email: pbottorf@nortel.com

Introduction:

The market for carrier-grade Ethernet equipment is growing as enterprises look to Ethernet services for more cost-effective bandwidth and service providers look to reduce their own network infrastructure costs to increase service profitability. Essential components for Ethernet equipment to support these carrier applications are: scalability, reliability, hard QoS / traffic management, service management and support for TDM services. To date, MPLS-based solutions have been the only technology option that solved all these requirements. Carrier Ethernet standards including the IEEE 802.1 Provider Backbone Bridging (802.1ah) and Connectivity Fault Management (802.1ag) standards are changing this landscape by providing new alternatives for metro networks that are based on Ethernet forwarding, simplicity and cost curves.

This paper introduces a new technology called Provider Backbone Transport (PBT) which can be employed in the service provider domain of a Provider Backbone Bridge Network(802.1ah) to allow configuration of engineered, resilient, SLA driven point-to-point Ethernet trunks. These PBT trunks allow carriers to engineer traffic managed circuits which may be monitored, along with the rest of the 802.1ah network, using 802.1ag protocols.

Provider Backbone Transport Overview:

In a standard Provider Backbone Bridged Network(802.1ah PBBN) traffic engineering is today limited as a consequence of the use of IEEE 802.1Q MSTP control plane protocols which control the population of the bridge filtering tables. The underlying 802.1Q/802.1ad/802.1ah bridge relays whoever don't have inherent characteristics which prevent full traffic engineering. Work in progress at IEEE on Shortest Path Bridging(802.1aq) is already working to improve link utilization by replacing the MSPT control plane with a shortest path spanning tree control plane. Provider Backbone Transport is a method for providing full traffic engineering of point-to-point paths in an 802.1ah network. To do this PBT replaces the MSPT control plane with either a management plane or external control plane and then populates the bridge filtering tables of the component 802.1ad and 802.1ah bridge relays by creating static filter table entries (see figure 1). The ability of PBT to create this external management or control plane is facilitated by 802.1ah because the B-DAs are all managed by the Provider and therefore can all be discovered and identified in the Provider's topology by the external management or control plane.

The external PBT management/control plane is responsible for maintaining and controlling all the topology information to support point to point unidirectional Ethernet Switch Paths (ESP) over the PBBN. The PBT topology can co-exists with the existing active MSTP or with the 802.1aq topology by allocating B-VIDs spaces to PBT, MSTP, or 802.1aq or can be stand alone. PBT therefore takes over control of a range of B-VIDs from the BCB and BEB bridges of the PBBN.

The PBT management/control plane forms a topology of B-DA rooted trees. For each <B-DA,B-VID> pair, configured in PBT, an independent tree is maintained. ESPs are routed by PBT along a tree selected by the B-VID to the destination B-DA. For a single B-DA the number of separate routing trees may be up to the number of PBT reserved B-VIDs. The B-VIDs may be reused for every B-DA, therefore the total number of routing trees maintained by the PBT management/control plane may be up to the number of B-DAs times the number of PBT reserved B-VIDs. The trees maintained by PBT for routing ESPs do not have to be spanning since they only require connectivity to all the source B-MACs which have ESPs to the specific B-DA. Each tree may connect to as many B-SA as desired with the only limits being implementation imposed table sizes. The PBT management/control plane may use any algorithm desired to selection the path for a routing tree providing complete route selection freedom. The PBT management/control plane also manages the bandwidth of all ESPs along each routing tree. For each B-SA which is part of a routing tree maintained by the PBT management/control plane PBT will maintain a routing tree which provides a co-routed reverse path from the B-DA to the B-SA. The B-VID used in this reverse ESP does not have to be the same one used for the forward ESP. The reverse ESP is used by CFM management to monitor the ESPs.

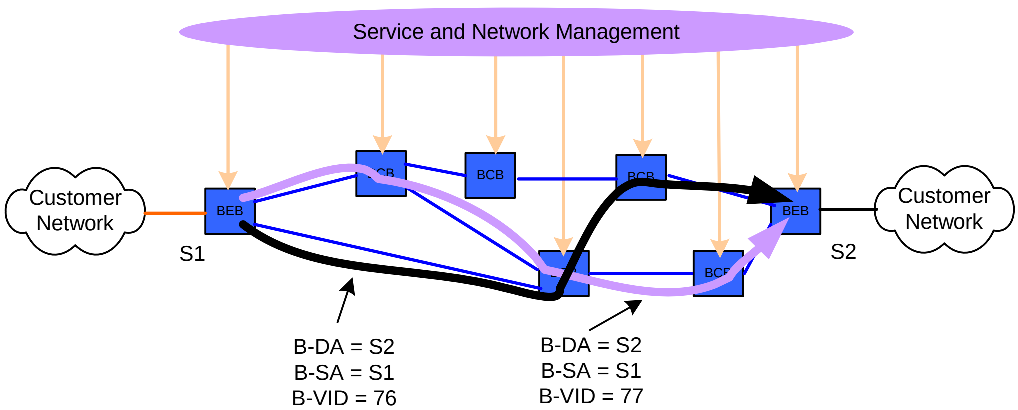

Figure 1—PBT Network

The relay functions used by PBT are standard bridge Individual VLAN Learning (IVL) which uses a B-VID/B-MAC tuple rather than a B-DA for filtering and forwarding decisions. When using IVL based forwarding it is possible to treat the combined <B-DA,B-VID> tuple as though it was a single 58 bit address where 12 bits are the B-VID and 46 bits are the B-DA. The IVL mode of bridge operation allows PBT to considers the B-VID part of the address as a path selector to the B-DA MAC rather than a B-VLAN ID. This allows up to 4094 unique routing trees to any single B-DA. Typically only a small number of B-VIDs are needed for PBT since it is normally not necessary to have even tens of alternate paths to a single destination. In figure 1 two paths are configured to reach S2. These two paths are separated by using a different B-VID in combination with the B-DA for a second path.

PBT requires no B-VID translation and operates on most 802.1ah unicast forwarding hardware. Even without B-VID translation the scaling properties are formidable. If the B-VID range delegated is the full 4094 possible values, then each B-MAC termination can sink 2^{12} routing trees, and the theoretical network maximum is 2^{58} ESPs (allowing for the multicast and local reserved bits in the MAC space). We do not need to impose completely new forwarding modes on Ethernet equipment or significantly re-specify the hardware and management to achieve traffic management.

To make PBT robust a few aspects of standard 802.1Q bridge forwarding must be changed:

1) Discontinuities in forwarding table configuration in the path will result in packets being flooded as "unknown". As there is no loop free topology for the delegated B-VID range, this will result in unbounded

flooding, looping and replication. For this reason flooding of packets with unknown destinations must be disabled for the designated B-VID range. Similarly broadcast and multicast traffic that would be flooded must be filtered at the ingress to the relay function.

2) B-MAC learning is not required, and may interfere with management/control population of the forwarding tables when combined with the potential of configuration errors. For this reason B-MAC learning is disabled for the delegated B-VID range.

3) This approach bypasses spanning tree for the delegated B-VID range, so spanning tree or multiple spanning trees are used only for the non-delegated (traditional operation) B-VID range.

4) When used in conjunction with a best effort spanning tree, CAC'd paths traffic requires a higher priority than best effort and in engineering the network, a reserve must be set aside for best effort traffic.

This approach has a number of useful properties:

The use of a global path identifier directly for forwarding with no translation is inherently more robust than alternatives. Any mis-configuration or forwarding table errors resulting in a deviation of a PDU from the prescribed path will self identify immediately. There is no possibility of collision with other identifier spaces that can mask the fault. This also suggests that no or minimal changes are required to existing CFM and Y.1731 constructs to successfully instrument Ethernet paths.

The ability to explicitly route and pin paths across the network can be combined with call admission control and 802.1Q class based queuing in order to provide per path QoS.

Example Ethernet Switched Paths

Figure 2 shows an example Provider Backbone Bridged Network running Provider Backbone Transport over the PBBN core. The PBBN B-VID space has been partitioned between MSTP and PBT with B-VIDs 7 and 8 allocated to PBT. PBT has taken over the port forwarding state machines in the Backbone Edge Bridges and Backbone Core Bridges for the allocated B-VID range and controls frame forwarding by adding static entries to the filtering databases of the switches within the PBBN core. MSTP operates normally in parallel to PBT on the other B-VIDs.

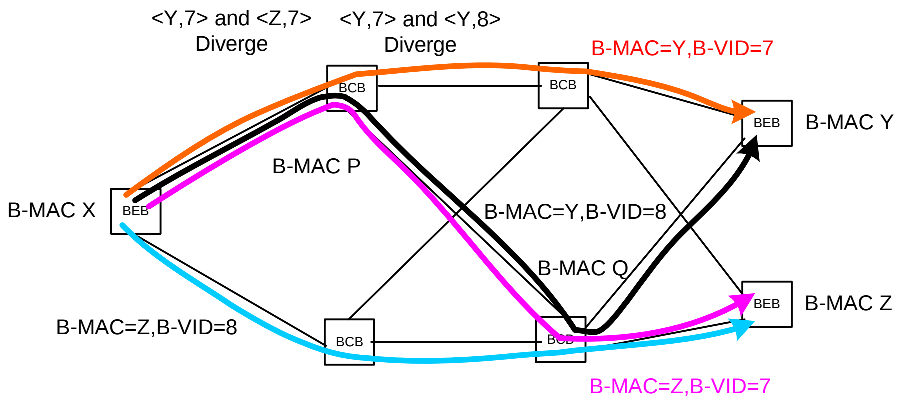

Figure 2—PBBN with B-VIDs 7 and 8 allocated to PBT operation

PBT operates in Indepentant VLAN Learning(IVL) mode where the combination of the B-MAC address and the B-VID can be thought of as a single address. The B-VIDs allocated to PBT are not treated as VLAN IDs, instead they are individual instance identifiers for one of a maximum of ‘n’, where ‘n’ is the number of allocated B-VIDs, possible routing trees to the destination B-MAC address. B-VIDs in the allocated range may be re-used for many routing trees within the PBBN as long as the <B-DA,B-MAC> tuple is unique. This PBT addressing arrangement results in a 58 bit globally unique destination address that may be shared by any number of B-SAs.

The example in figure 2 illustrates the complete route freedom of configured forwarding in IVL bridges. In the example a total of 4 ESPs use 2 B-VIDs to forward traffic to 2 B-MAC terminations. At switch 'P' above, despite collisions in both the B-MAC and the B-VID space, forwarding properly resolves because the switch uses the B-MAC and B-VID together to establish route uniqueness. At node P the red and purple ESPs diverge even though they have the same B-VID because they are addressed to different B-MACs. Likewise at node P the red and black ESPs diverge even though they have the same B-MACs because they have different route selectors (B-VIDs).

Summary:

Provider Backbone Transport (PBT) is a technique for providing traffic and network engineering of point-to-point backbone paths over 802.1ah networks. The technique operates by replacing the MSTP or 802.1aq control plane with a PBT management/control plane which maintains a set of PBT routing trees, each rooted at a B-DA and selected by a B-VID. The routing trees are used to route Ethernet Switched Paths within the PBBN. The PBT system provides a carrier the freedom in route selection over and 802.1ah backbone necessary for network and traffic engineering of a carrier network.