Hardware choices for MPLS and backbone bridging

Martin Nuss, PhD, Vitesse  - March 17, 2011

The traditional model of evolving hardware support for more complex routing in carrier-based network equipment assumed that smaller silicon features, more optimized design techniques, and stabilized high-layer protocols would converge to allow single-chip switch devices to climb up the protocol stack. Switches that once supported only Layer 1 PHY (physical) and Layer 2 data-link functions would grow over time to include Layer 3 network functions, Layer 4 transport duties, and a mix of higher-layer presentation and application functions to support duties such as deep-packet inspection.

In a broad-brush approach, this vision still holds. The arrival of MPLS (multiprotocol label switching) to provide IP (Internet Protocol)-switching capabilities in scaled hierarchies of routers played an important role in simplifying routing for silicon. In the middle of the last decade, however, some equipment providers and carriers expressed interest in PBB (provider backbone bridging), a simple Layer 2 protocol, throwing a wrench into the just-simplified works. In some sense, this move was a replication in the public network of the bifurcation in enterprise networks between routing and bridging. Consequently, you could look for lessons from the evolution of bridging that took place in the early 1990s. As you will see, however, you cannot assess the current “debate” between PBB and MPLS as a simple replay of the LAN (local-area-network) bridging-routing dichotomy, taken to the WAN (wide-area-network) domain.

In LAN-based bridging, packets using similar protocols transfer from one subnet to another. Ethernet’s shift from shared hub and bus topologies to switched hub-and-spoke topologies in the early 1990s streamlined and simplified bridging. As Layer 2 switching displaced all shared-Ethernet LANs by the end of the century, silicon providers were able to standardize switch designs and integrate multiple ports on one switch device. At first, vendors offered quad and octal switches without PHY or full MAC (medium-access-control) functions on the switch chip. As semiconductor providers turned to deep-submicron processes, however, switch chips scaled to 24 ports and more, even as they integrated MAC and PHY functions in the central switch chip.

Switch chips for the LAN could improve along another dimension, as well. Centralized routers and edge-based access devices, which had to interface with protocols other than Ethernet, added management intelligence and MPLS support to their core switch chips, thus making these devices Layer 3 switches. This process began with ASICs and FPGAs because many early routers relied on proprietary interfaces to WANs or dissimilar LANs. The victory of Ethernet and TCP (Transmission Control Protocol)/IP over ATM (asynchronous transfer mode), however, combined with the standardization of edge-routing functions in the enterprise and allowed the Layer 3 switch to become an ASSP (application-specific standard product). Easier integration of glue logic allowed vendors to integrate the Layer 3 functions into reasonably priced switches, but a Layer 3 switch always carried a price premium over its Layer 2 equivalent. The semiconductor-developer world
adopted a slogan that originated in the system-level-network-equipment world: “Bridge—or switch at Layer 2—when you can and route when you must.” To a certain extent, that same philosophy applies in today’s service-provider WAN, although the debate between PBB and MPLS is not a simple replication of the bridging debate of the past.

The newest version of PBB, PBB-TE (PBB transport engineering), suffers from the fact that the word “bridging” remains in the technology description. The type of backbone bridging that 21st-century multigigabit backbones use is not your father’s bridging, and its complexity stems from more than just the order-of-magnitude improvements in speed.

**MPLS and PBB-TE: What’s the point?**

A natural tension exists between a connection-oriented deterministic world, in which service guarantees are a natural response to the problem of deterministic delays, and the connectionless packet world of the Internet. This tension forms the basis of the transport technologies that MPLS and PBB-TE defined. Although the early TCP/IP networks had no true nature-of-service guarantees, in practice carriers did offer such assurances in a de facto manner at higher network layers, to give a connectionless, packet-switched topology the same robustness as SONET (synchronous optical network)/SDH (synchronous digital hierarchy).

Before considering the relative merits of either PBB-TE or MPLS, it is fair to ask whether users still need any form of connection-oriented emulation, now that advances in deepsubmicron silicon allow significant improvements in speed of IP-header look-ups. As TCP/IP began overtaking ATM switching in the late 1990s, developers looked for some way to provide some of the features of ATM’s VCs (virtual circuits) to an IP-centric world. Members of the IETF (Internet Engineering Task Force) who worked on MPLS recognized that ATM’s advantages in service guarantees and VC visibility were too important to abandon in a TCP/IP-centric world and needed to be available in a connectionless domain.

The motivation for the MPLS proposal at that time was the latency in performing full IP longest-prefix matches at each hop in a routed network. By attaching an MPLS shim header to a packet at ingress, the network could identify packets and forward them in a switched network with speeds that could easily achieve SONET/SDH-like resiliency targets of less than 50 msec. Since MPLS first became a standard, silicon advances have allowed routed networks without MPLS to achieve high forwarding speeds. Nevertheless, both MPLS and PBB-TE have achieved justification outside performance parameters by offering a means of creating point-to-point “tunnels” that provide an easy means of offering services such as pseudowire and VPNs (virtual private networks).

The full G.8110 MPLS allows but does not mandate the use of routing and signaling protocols for topology and resource discovery. A centralized network-management system can fully provision MPLS LSPs (label-switched paths) without the use of any localized routing or signaling protocols. To further simplify the use of MPLS, the IETF identified the MPLS-TP (transport-profile) subset, which assumes the use of a connection-oriented packet transport employing a centralized static provisioning model, eliminating all local routing and signaling protocols. This step eliminates many protocols that developers built for MPLS, such as ECMP (equal-cost multiple path) and LSP merge. The MPLS-TP model allows the use of multiple client layers, including MPLS, native Ethernet, and TDM (time-division multiplexing).

PBB-TE and its predecessor, PBB, grew out of IEEE 802.1ad’s QinQ effort, which allows service providers to create VLANs (virtual LANs) with a carrier-provided tag. Although a VPN employing MPLS hides a customer’s MAC address, the original QinQ did not hide this address, requiring the
entire network core to learn clients’ MAC addresses. PBB in its original form, 802.1ah, dispensed with the spanning-tree protocols that 802.1ad used and eliminated the broadcasting of unknown MACs. Because of the way it defined MACs, its new header, MAC-in-MAC, referred to the complete re-encapsulation of a customer frame at the edge of the public network, hiding the customer address from the service-provider core.

Although both PBB-TE and MPLS can offer VPLS (virtual-private-LAN service), PBB offers these services as a bridged Ethernet tunnel. MPLS offers protected LSPs that users can centrally engineer and protect. The IEEE’s incentive to move to 802.1Qay, or PBB-TE, was to overcome some of the perceived limitations of standard PBB. In the new PBB-TE standard, MAC learning and spanning-tree protocols are turned off for all backbone VLANs. Either a centralized network-management system or a GMPLS (generalized MPLS) control plane populates bridge tables with a static forwarding entry.

When the founders of the MEF (Metro Ethernet Forum) first met in 2001, their goal was to develop carrier service types, using strict Layer 2 systems when possible. Consequently, the MEF was closely associated with bridging and QinQ VLANs as it developed E-Line, E-LAN, and E-Tree (point-to-multipoint) service offerings (Figure 1). Although the MEF does not take an official position on pure Layer 2 versus MPLS Layer 2.5 approaches and emphasizes the cost-effectiveness of pure bridging links for smaller networks, the forum optimized the service profiles for MPLS-based networks. Because the MEF advocates advanced OAM (operations/administration/maintenance) and performance-management tools, it supports the convergence of MPLS and emerging IEEE standards, such as 1588 (synchronous Ethernet) and 802.1ag (Ethernet fault management).

In theory, PBB-TE can offer most of the features of CE (carrier-Ethernet) service creation and tunneling, using a simpler hardware implementation employing Layer 2 switches and small per-port memories. PBB-TE is still at least one to two years from finalization, however. Even now, the cost delta is shrinking between PBB-TE and transport MPLS in multiport Ethernet chips. Does it continue to make sense to use PBB-TE in cheaper networks closer to the edge and preserve MPLS for metro core and aggregation duties, or does end-to-end MPLS make sense from an SOC (system-on-chip)
MPLS has some unique functions that may be critical for some carriers. For example, it uses a label stack with a 3-bit traffic-class field that defines both the QOS (quality-of-service) level and explicit congestion notification. Careful implementation of 802.1Qay can offer some QOS features, but MPLS carries a natural advantage in defining classes of traffic for preferred delivery options in service-level agreements. Fast-reroute features in MPLS more easily provide the ability to hit SONET/SDH-like less-than-50-msec protection switching speeds than many Layer 2 alternatives can offer.

Market-analysis companies, including IDC, have produced studies that indicate that upgrading Ethernet switches for PBB-TE support will add 33% to real estate and overall design costs. In contrast, the cost of adding Layer 3 functions for MPLS can vary tremendously based on factors such as numbers of ports. Therefore, no transport-tunneling method is cost-free. Because PBB-TE eventually will migrate to GMPLS as a control plane, the issue is where and why a low-cost bridging approach employing native Ethernet transport might make sense over comprehensive MPLS support.

Even if you assume that PBB-TE in its finalized form was technically antiquated upon completion of the standard, you can scarcely call the draft technology obsolete in the market of 2011. At the beginning of 2010, PacketExchange, an IP-specialized service provider using MPLS transport, acquired carrier Mzima for its PBB-TE expertise. Service providers standardizing on TCP/IP at layers 3 and 4 and Ethernet as a Layer 2 packet-framing method are coming to a common conclusion on the coexistence of MPLS and PBB-TE. As MPLS becomes more cost-effective to consider outside the network core, it may find more use than does PBB-TE. PBB-TE bridging technology, however, likely will have a place at the edge of the public network for years to come.

It makes sense for service providers to plan for MPLS-dominated networks. Thus, over time, Ethernet-chip developers must design for OEM products that emphasize greater MPLS features. The IEEE and MEF both see Layer 2 bridging as a constrained alternative to the multiple layers of transport guarantees that MPLS offers as a transport method. A mature Ethernet technology in metropolitan carrier networks should include such features as protection switching and distributed timing. In such networks, CE topologies can be natural adjuncts to the sort of IP/MPLS networking strategy that operates in both service-provider networks and large-scale enterprises. Silicon providers must accordingly offer comprehensive suites of switching, MAC, PHY, and coprocessor chips, with embedded support for PBB-TE and transport MPLS, allowing the OEM and, thus, carrier customers to carefully design the transport for the task at hand.

PBB-TE may be less strategic over time when planning for global backbone networks that large carriers manage. It can be ideal, however, for smaller regional carriers upgrading TDM-based metropolitan subnetworks or defining a new “greenfield”-network topology—that is, designing and implementing a network from the ground up. PBB-TE plays its most explicit role in metro networks largely employing full-chassis fault-resilient Ethernet switches. Today’s Ethernet switches that seek to take on a carrier-transport role for a metropolitan-service-provider network must show fault tolerance, determinism, and sufficiently accurate network timing to allow transport of isochronous traffic.

Provided that a carrier network is well-characterized, PBB-TE can fulfill carrier-network QOS goals that have risen to a level that no one expected when developers recently designed some of the first VLAN-tagging methods. Networks can realize OAM functions, for example, through the use of the IEEE’s 802.1ag standard for connectivity fault management. The network can establish a path-protection mechanism similar to that of SONET/SDH by creating separate backbone-VLAN
Identifiers for a work channel and a protection channel. With such extensions, the 10-Gbps CE switches now on the market look more like fault-resilient digital cross-connects than like enterprise Layer 2 switches.

Although PBB-TE can offer sophisticated OAM and performance-management functions, many emerging ITU (International Telecommunications Union) standards, such as Y.1731 and G.8031, work better in an environment extending MPLS to the network edge. A supporter of PBB-TE might insist that Y.1731’s more expansive OAM goals depend more on new dedicated PHY capabilities at the semiconductor or subsystem level than on a shift from bridging to Layer 2.5 label-switching. To a certain extent, that belief is true. Under MPLS, however, analysis of frame delay, locked conditions, packet loss, and the like can take place in an environment of broader traffic encapsulation and VC creation than is possible in a strict PBB-TE environment.

For example, G.8031 defines a signaling protocol that turns a VLAN link into an automatic-protection switching path, so designers could add the G.8031 approach to VLAN tags in duplicated backbone-VLAN identifiers that PBB-TE defines. The wealth of unidirectional and bidirectional protection-switching methods in G.8031 maps better into an MPLS topology, however, and both transport methods ultimately want to move to a GMPLS control plane, eliminating the differences between the two camps.

Recent advances in edge routing indicate that full Layer 3 control-plane approaches can move closer to the edge, although an access approach in which PBB-TE makes the most sense will also likely remain. The critical issue is whether MPLS serves strictly in a transport dimension or as a full control plane. The scalability of the control plane becomes the critical issue as MPLS moves to the access platform. The preprovisioning of physical or virtual circuits often makes more sense than extending MPLS control planes over large topologies.

Specialized metro service providers considering PBB-TE should expect equipment OEMs to upgrade simple bridged-Ethernet topologies in the public network with hardware and firmware supporting fault resiliency and protection-switching features, many of which 802.1Qay spells out. As a result, PBB-TE will continue for many years to come to play a role in implementing MEF services. It is too soon to bet against MPLS. As semiconductor features shrink and SOC designs become larger and more complex for a smaller unit price delta over simpler chips, the case for end-to-end MPLS transport seems to have a broader rationale than that for PBB-TE.