Ethernet Data Plane Evolution for Provider Networks

Don Fedyk and David Allan, Nortel

ABSTRACT

This article provides an overview of the evolution of the Ethernet data plane. In particular, it focuses on the emergence of features that have made Ethernet an attractive infrastructure technology option for carriers and network providers. These include the development of data plane maintenance protocols (OAM), and scaling enhancements, initially in the form of VLANs, then VLAN stacking (802.1ad), and more recently complete recursion of headers (802.1ah). The complete recursion of headers has led to the ability to decouple the infrastructure data plane from traditional bridging behavior while preserving other desirable attributes, leading to new approaches to operating Ethernet networks such as PBB-TE.

OUTLINE

This article follows the evolution of Ethernet's data plane history up the recent developments. Ethernet and IP are contrasted with some detail on Ethernet services. Then we cover bridging with the introduction of VLANs, provider bridging (PB), and the latest provider backbone bridging (PBB). Next, the new operations, administration, and maintenance (OAM) features are covered, and we introduce PBB-traffic engineering (PBB-TE) that leverages PBB and provides control plane independence. Finally, we wrap up with the combination of PBB and PBB-TE, and provide conclusions.

LEVERAGING ETHERNET'S DATA PLANE HISTORY

Ethernet has a long history. From simple beginnings in 1974, a ubiquitous interconnect technology has emerged. Ethernet specifications in the IEEE 802.3 working group define the Ethernet for local area networks (LANs) and the Ethernet frame structure. In the IEEE 802.1 working group they define bridging, the forwarding process for connecting Ethernet LANs. Originally designed for simple low-cost access and LANs, Ethernet has stood the test of time.

This article focuses on the data plane aspects of Ethernet bridging. Although bridging has control plane aspects associated with spanning trees, the main focus is the attributes of the data plane.

The last few years has seen renewed interest in both extending and reinventing Ethernet. Encompassed in Ethernet standards are both link and physical layer definitions. While Ethernet can be adapted onto other media, it is emerging as being completely self-contained, offering packet framing, switching, and integrity services to client layers.

Today, Ethernet is one of the highest-speed packet technologies in part due to simplicity and subsequent commoditization. Ethernet at the physical layer has been on the leading edge of increase in packet data rates. Ethernet has typically increased the highest data rate by an order of magnitude every three to four years, and has gone from 10 Mbs/s to 10 Gbs/s today, with 40 and 100 Gbs/s on the horizon.

ETHERNET AND IP

Ethernet frames carry practically any protocol. IP, of course, is the most dominant data protocol today, and IP has been well adapted to ride over Ethernet with some 750 supporting Internet Engineering Task Force (IETF) Request for Comments (RFC) references. Ethernet, being a variable length frame based technology, efficiently carries IP packets. Ethernet has acted as the first level of aggregation for IP networks since the start of the Internet, and a suite of protocols exists to facilitate Ethernet attachment of routing-unaware hosts to the Internet. The dominant examples are Address Resolution Protocol (ARP), Dynamic Host Configuration Protocol (DHCP), and Internet Control Message Protocol (ICMP). The broadcast capabilities of Ethernet simplify many of the functions of the first mile such as resiliency, auto-discovery, and a first level of address aggregation.

However, the increased use of Ethernet as infrastructure changes the mix of attributes that are desirable, and provides motivations for moving away from the less scalable and undesirable aspects of learning and broadcast behavior.

IP provides global any-to-any connectivity with the proviso that Internet providers effectively peer with their customers as well as with other providers. The Internet is a single large communi-

0163-6804/08/$25.00 © 2008 IEEE

ty of interest. Ethernet, on the other hand, provides connectivity between constrained or provisioned sets of network interfaces. As we will see in this article, one of the goals of carrier Ethernet evolution is to allow global Ethernet connectivity between provisioned sets of customer interfaces. The objective with Ethernet is different from the use of IP to underpin the Internet, because it addresses the problem of connecting large numbers of smaller communities of interest.

While the networking paradigm differs between IP and Ethernet, there is a significant amount of overlap in functionality between the two data planes. For functions like advanced queuing techniques, service classes, security, encryption, and congestion management, there is sufficient overlap that Ethernet networks can satisfy the requirements of transporting IP.

While IP routers have added functionality in recent years, Ethernet bridges have been less active and more rigorous in the specification of new features that have been adopted. The simple networking paradigm of bridging, a backward compatible architecture and measured incremental change has allowed Ethernet to leverage and maintain a high degree of commoditization.

Ethernet Services Spectrum

While the drivers for Ethernet evolution are global in nature, the whole technology — from simple LANs to carrier services — is undergoing subtle but revolutionary change. The Metro Ethernet Forum (MEF) lists five attributes that define carrier class service:

- Standardized services
- Scalability
- Reliability
- Quality of service
- Service management

These attributes vary in their applicability across the range from home networks to global networks. The evolution of Ethernet is as much about providing high definition audio and video in the living room as it is about providing the highest-speed packet interfaces in the industry. The evolution of carrier Ethernet is also about providing enterprises with a suitable carrier interface when the enterprise controls their own network. While this article focuses on provider backbone aspects, it is important to realize that the changes to Ethernet are more broadly based.

Standardized Services

Ethernet services are determined by two aspects: the definition of a standard interface with its capability, and the definition of a switching capability to support these services. These two aspects are linked by common attributes, but are subtly different. For example, the use of the virtual LAN (VLAN) identifier has a simple meaning when defined as an interface, but a more encompassing role in bridging as we will see.

The evolution of Ethernet has been driven by the need to provide a set of standardized Ethernet services that can be simply defined and easily deployed. These services transport Ethernet and whatever Ethernet is carrying. The MEF defines three basic types of Ethernet virtual circuit services, E-LINE, E-LAN and E-TREE, as a reference for three types of common Ethernet services.

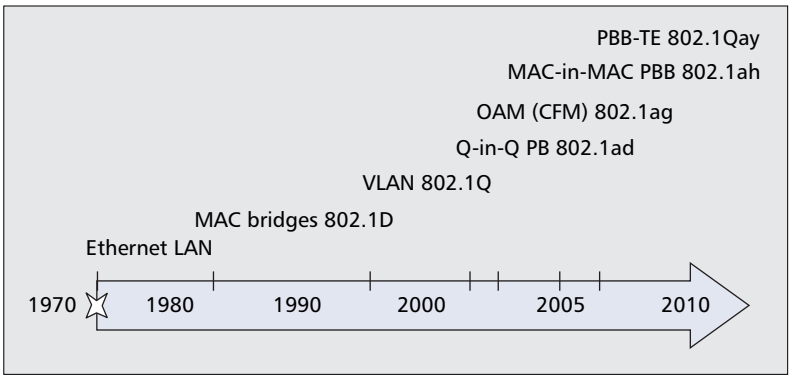

Figure 1. *Timeline for significant Ethernet enhancements.*

E-LINE is a simple and basic point-to-point Ethernet circuit. These are often called virtual private lines, and can carry any type of packet traffic for which an Ethertype is defined.

E-LAN is an Ethernet LAN service where unicast, multicast, and support for services like IP are all important. E-LAN services involve the provider networks by requiring them to support the basic LAN functions of learning, unicast, and multicast just as a private LAN would. The ability to support any-to-any multicast in these types of services is a challenge in most networking technologies. This native ability to support multicast is Ethernet's strength.

E-TREE is a reduced form of E-LAN service where a dedicated source is able to multicast to all destinations, but the destinations are only allowed to respond to the source. One typical application of E-TREE service is the broadcast of video to residential customers or providing layer 2 isolation between customers in an access/aggregation network.

To summarize, the three services can be characterized as connectivity primitives: E-LINE is point-to-point, E-LAN is multipoint, and E-TREE is client-server.

Ethernet Bridging

Ethernet began as a single shared medium. There has been a progressive evolution of interconnect and forwarding, with repeaters first evolving to hubs, and then evolving to bridges with progressive increases in scale and efficiency (Fig. 1).

Ethernet bridging, often termed *transparent bridging*, is the mechanism responsible for relaying and replicating Ethernet frames within an Ethernet network. Bridging typically requires the setup of a spanning tree that can reach all other bridges within the LAN. The role of the spanning tree is to provide a loop-free topology within which broadcast, multicast, and flooding and learning can operate. A single minimum spanning tree was used initially and to this day remains a common deployment scenario. This choice of architecture allowed both unicast and multicast traffic services simultaneously in any Ethernet bridged network.

Ethernet traffic flow is bidirectional and symmetric; the forward and reverse paths between any two points in a stable network are exactly congruent. The moment a frame is switched onto a spanning tree, it must follow the tree to the destination. In a stable tree, packet order

The physical reach limitation of copper wire Ethernet constrained the applicability of Ethernet in the provider space. Once optical Ethernet interfaces emerged with vastly increased reach, providers were able to apply Ethernet technology directly into their networks.

and predictable delivery are ensured by following the tree. This has some very nice properties for transport networks.

A key aspect of bridging is the concept of low or zero configuration for essentially plug-and-play operation. This is achieved by unique allocation of medium access control (MAC) addresses (typically global allocation) and leveraging the broadcast capability inherent to the spanning tree to learn unicast addresses within the domain of the network. Bridges learn the MAC address of active endpoints by observing where packets come from and updating the forwarding tables accordingly. When a bridge receives a frame for an unknown destination it is flooded, with the expectation that subsequent return traffic will be observed by the learning process and fill in the unknown forwarding information. In many ways this is a backward compatible behavior inherited from the early days of Ethernet shared media. The returning frames are learned and cached for a period of time. This enables simple and efficient unicast forwarding with low overhead, with a graceful mechanism for coping with forwarding table exhaustion by trading off bandwidth efficiency.

VIRTUAL LANS (IEEE 802.1 Q)

In recent years the widespread use of Ethernet found its way into carrier networks for offering point-to-point and VLAN services. The simplicity and cost efficiency of Ethernet equipment make it very attractive, but the demanding requirements of carrier networks require the simple capabilities of original Ethernet to be extended.

VLANs [1] were introduced in the mid-1990s to logically constrain a bridged topology to an arbitrary subset of the physical network. The VLAN identifier (VID) is a field in the Ethernet header that has 12 bits, allowing 4096 VLANs. VLAN introduced several important concepts as mentioned earlier. There is the concept of an active VLAN topology for the set of bridge ports that share the VLAN. Then there is the concept of a service provided by the VLAN to the interface set of ports. This is the reason that VLANs on an interface or a point-to-point circuit behave simply as channel identifiers for multiplexing services.

In Ethernet switching, the active topology concept of the VLAN is useful to limit the broadcast domain and control domain of the LAN. By creating a VLAN and assigning ports to the VLAN, a community of interest is created. This community is a closed user group, typically accessing a common service network. Many VLANs may operate in parallel, and backward compatibility allows shared VLAN operation as well.

As VLANs were being developed, the notions of service class and quality of service (QoS) were also being addressed. After a number of solutions were explored, a tag header was created to extend the Ethernet frame. This Q-Tag header carries the VLAN ID, and also contains service priority. This is a direct continuation of the design principle of Ethernet; a frame explicitly identifies the queuing discipline to be applied to it.

PROVIDER BRIDGES (IEEE 802.1AD)

The physical reach limitation of copper wire Ethernet constrained the applicability of Ethernet in the provider space. Once optical Ethernet interfaces emerged with vastly increased reach, providers were able to apply Ethernet technology directly into their networks. This led to a conflict for VLAN use, as this capability was exploited by both service providers and enterprise customers.

Service providers began offering layer 2 services between customer sites. These layer 2 services were viewed by the customer as shared media, a LAN or VLAN. The actual implementation of the provider network was initially based other technologies such as asynchronous transfer mode LAN emulation (ATM LANE) and later virtual private LAN service (VPLS). Customers also used their own VLANs for QoS and simplified layer 2 management. When a provider offered a VLAN service to the customer, the provider would have to honor the customer's VLAN properties.

These situations led to a number of solutions being proposed to preserve the customer characteristics. The solution that gained widespread acceptance was to stack the Q-TAG, creating the so called Q-in-Q header named after the Q-TAG (Fig. 2).

Stacking of VLANs to distinguish a customer VLAN (C-VLAN) from a service VLAN (S-VLAN) allows a service network (typically a provider) to administer their own VLAN space while carrying another client network's (typically a customer) VLANs transparently.

There are two important aspects here. First, there is a mechanism to stack customer service VLANs in provider VLANs. Second, there is the need for some technology underlying this to support the provider VLAN. When the technology is a bridged network, the multicast and unicast services are supported natively by the provider infrastructure, and there is no need for LAN emulation.

In essence, with PB, Ethernet had evolved to address the provider market natively. While PB solved the problem of multiple administration of the VLAN space, two problems that limit its scalability remained. First, while the PB system cleverly exploited current hardware, it gave the provider only 4096 service VLANs, limiting the network to 4096 service instances. The second issue with PB is that while the C-VLAN was hidden or encapsulated from the provider domain, the MAC addressing was still visible within the S-VLAN space, requiring provider bridges to learn and forward customer addresses.

PROVIDER BACKBONE BRIDGES (IEEE 802.1AH)

During the project approval for provider bridges (IEEE 802.1ad), the concept of complete encapsulation was debated. Provider backbone bridging (IEEE 802.1ah) [2] is the culmination of this evolution, allowing full encapsulation of the customer functions of topology and service identifying frames (Fig. 2). PBB utilizes an 802.1Q standard header and an S-VLAN Ethertype but it separates the backbone VLAN (B-VLAN) into a VLAN plus a service identifier (I-SID), see Fig. 2. This is important since the number of VLAN topologies is typically a scaling constraint for Ethernet. By allowing any subset of the topology for services, the services scale independent of topology, and the B-VID is delegated the role of engineering the network.

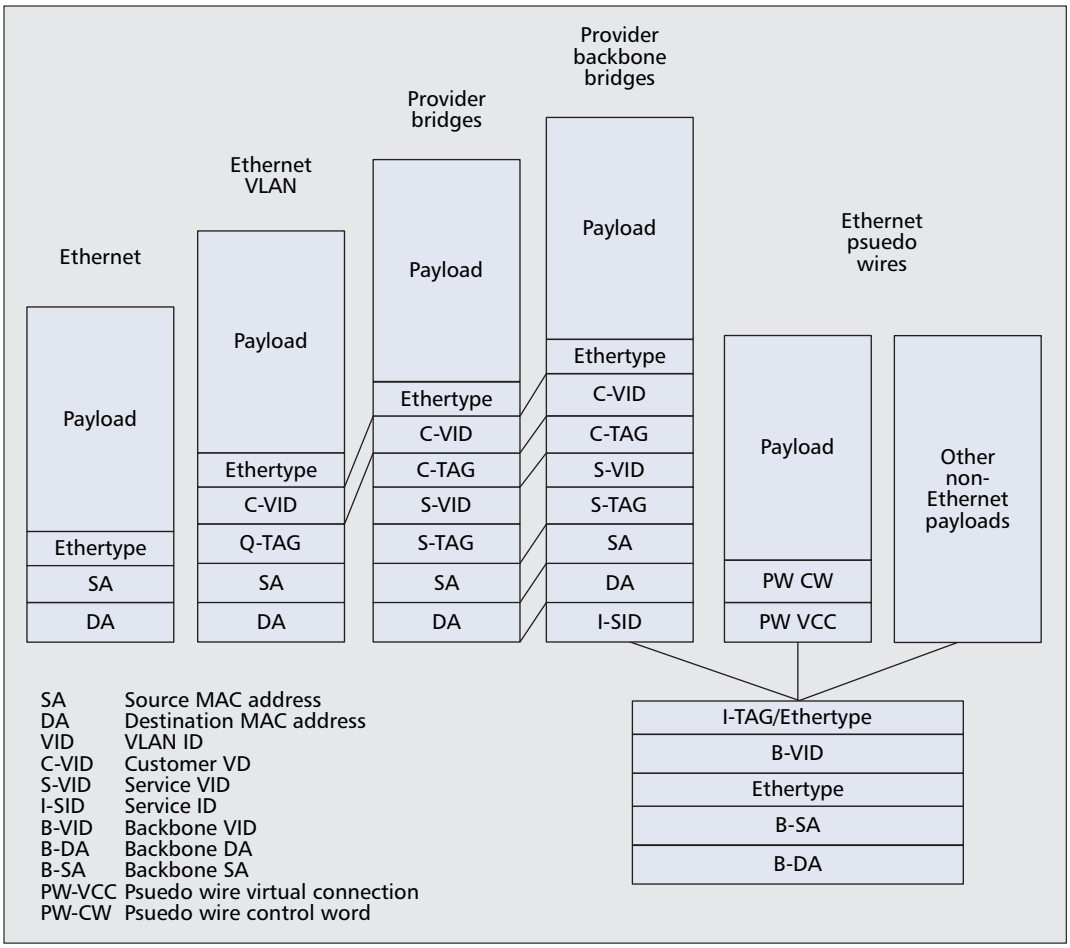

Figure 2. *Evolution of the Ethernet header.*

During the standardization of PBB much work went into defining the new I-Tag for the service instance identifier. The separation of the service from the VLAN is new and it created a number of problems for the OAM.

The complete encapsulation provides for a comprehensive customer-provider demarcation point. The service provider network only transports frames in a provider frame format containing provider administered identifiers. This allows the service provider to separate the topologies used by different customers or aggregations of customers, by controlling the mapping of I-SIDs to different B-VLANs. Many customers can be supported on a single B-VLAN.

This service identifier thus allows for a greater degree of flexibility in managing services by allowing independence from the topology.

The other advantage of encapsulation is that customer addresses and customer MAC learning are isolated to the provider edge with the adaptation function providing the mapping between the customer MAC space and the provider MAC space. As the number of backbone edge bridges are orders of magnitude lower than the number of customer MAC endpoints supported by the PBBN, the overall scalability of bridging increases by a corresponding amount. Scalability is now global as interconnected sets of provisioned addresses are merely edge-based. Also, the provider edge can now be instrumented independent of the customer addresses. Separation allows the control plane functions of the carrier to be independent of the customer.

PBB networks are backward compatible with PB networks (Fig. 3). A PBB network may encapsulate a PB network that encapsulates a customer network. And PBB networks can be encapsulated in other PBB networks or peered with other PBB networks. Hierarchy provides scalability by aggregating customer networks. PBB achieves this with a bridged core that can always support the multicast and unicast model natively.

During the standardization of PBB, much work went into defining the new I-Tag for the service instance identifier. The separation of the service from the VLAN is new, and it created a number of problems for OAM (see the next section). PBB OAM has to be extended to work not only within a single provider backbone, but also to ensure the model works for multiple peered backbones. Although peered PBB networks are less likely than standalone PBB networks initially, in standards it is important to run the technology through various scenarios to make sure it can address future deployments.

The I-SID is a service identifier that is unique and consistent within a provider network. An I-SID uniquely identifies a virtual community of interest that is implemented as a virtual broadcast domain over which customer-transparent bridging can operate. For a service the I-SID identifies the grouping. Figure 4 illustrates the PBB I-SID and B-VLAN separation.

For point-to-point services, the function of the I-SID and multiplexer label is remarkably similar, but the multicast and unicast nature of

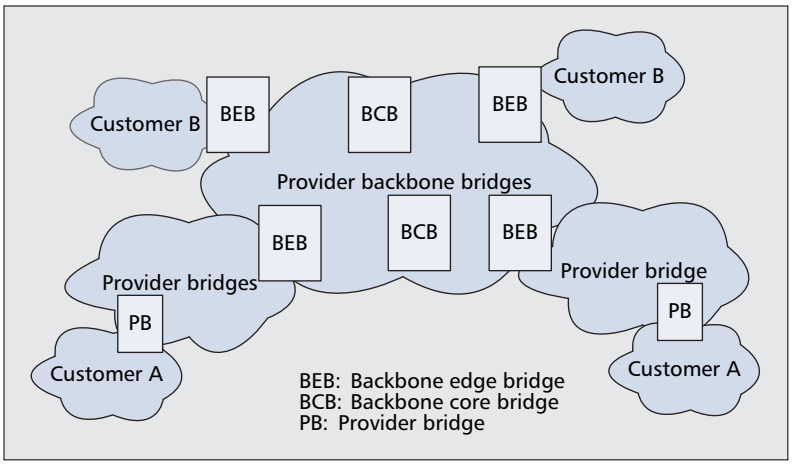

Figure 3. *Provider backbone bridge hierarchy.*

Ethernet gives the I-SID a consistent meaning in both scenarios. It does mean that a mechanism is required to ensure network-wide I-SID consistency, a concept that has been used in similar technologies. However, changing the paradigm for those I-SID properties that are inherited from the S-VLAN to use a local label would change Ethernet architecture fundamentally.

OAM (IEEE 802.1AG, Y.1731)

Another response to carrier and customer requirements has been the development of OAM (802.1ag [3], Y.1731 [4]) capabilities to comprehensively instrument the data plane. When we refer to OAM in this article, we are referring to the data plane OAM protocols that support network operations and fault/alarm and performance management. This is distinguished from the craft and EMS OAM interfaces that exist to support configuration and gathering of network and service statistics.

Ethernet OAM was originally forged in the IEEE 802.3 Ethernet working group. Arising out of the PB project was a need to supply OAM for a number of functions.

A requirement on the OAM procedures for bridging was that they be solely dependent on the data plane. This has several benefits: OAM works regardless of the control plane type, or even if one is used; OAM follows the true data path more closely and can be tunneled through transparently.

Ethernet has a number of architectural properties that make it amenable to the application of data plane OAM as a closed system. For example, bidirectional congruency is leveraged for loopback and fault management and performance management procedures.

The OAM functionality that has emerged from the IEEE and International Telecommunication Union — Telecommunication Standardization Sector (ITU-T) includes a suite of fault management, alarm management, and performance monitoring tools. These are defined as a protocol suite and exercise a range of functionalities in the data path, typically by using the same frame formats and forwarding procedures as normal traffic, with the OAM flows being distinguished from regular traffic by the receiver using the Ethertype.

The IEEE tools provide a basic fault management suite that includes:

- CCM — connectivity check message: a multicast heartbeat using a reserved address
- ETH-LB and LT, loopback and link trace, analogous to IP's ping and trace route

The ITU-T tools both augment the fault management set by defining unicast variations of the fault management tools, but add performance and alarm management transactions such as:

- LM — loss measurement
- AIS — alarm inhibit signal
- RDI — reverse defect indication

None of these have direct analogies in the IP world but are well known to designers of L1/L2 OAM.

Provider Backbone Bridging — Traffic Engineering (IEEE 802.1Qay)

PBB-TE [5] is the IEEE project name for an initiative initially brought to market under the broader banner of provider backbone transport or PBT. PBB-TE built on the separation of Ethernet as a service from Ethernet as infrastructure pioneered by PBB.

The ability to supply simple point-to-point private connectivity primitives leveraging Ethernet was an intersection of the PBB encapsulation and the static nature of provider networks. PBB-TE decouples the Ethernet control and data planes by turning off flooding of unknowns and source learning, and disabling the active topology. Although the initial focus of this work was on point-to-point circuits, the PBB-TE paradigm of explicit configuration of global addresses is equally applicable to construction of multicast trees, so leveraging the full capabilities of the native Ethernet data plane. Fundamentally PBB-TE does not alter the Ethernet forwarding paradigm and reuses the existing administration of identifiers. PBB-TE still uses destination-based forwarding based on VID and MAC DA; therefore, the data plane objects can have an order *n* relationship with the number of destinations.

In essence, PBB-TE adds an explicit pinned path mechanism to a PBB network (or, as it is called in standards, a PBB traffic engineered connection). Thus, the native Ethernet behavior used by PBB is still available, but PBB-TE partitions off some VLANs exclusively for point-to-point services (E-LINE) and managed point to multipoint services. A subtle aspect is now coming into play. While multiple VLANs were allowed, the requirement to co-exist with Spanning tree was new, and provided a mechanism for backwards compatibility with existing deployments. PBB-TE partitions off a set of VLANs for PBB-TE connections, using the VLAN set as a forwarding mode selector.

PBB-TE requires that forwarding tables be programmed by provisioning or by a control plane, removing the requirement for learning in the data plane. The loop-free requirement for

bridging is satisfied on a per-connection basis in the management or control plane, allowing traffic engineering of connections along arbitrary paths, completely divorced from any simple tree.

In order to create resilient services a way was needed to protect PBB-TE services. The model currently being proposed uses ideas very similar to those found in protection switching currently used for time-division multiplexing (TDM) networks. The most fundamental of these is to completely delegate protection to simple autonomous data plane mechanisms, and require the management or control plane only to configure protection mechanisms, not execute them. In this way preprovisioned disjoint backup connections can protect connections in a 1:*n* or 1:1 fashion.

PBB-TE can be viewed as introducing a domain-wide label: the VID + MAC DA. In this aspect it behaves as a globally allocated label. Generalized multiprotocol label switching (GMPLS) is one technology that has been applied to optical switching, where *labels* are wavelengths with constraints, and thus conceptually closely analogous to the PBB-TE global label. GMPLS has been proposed to control PBB-TE labels [6].

PBB and PBB-TE Combination

Over the course of the last five years, the goal of producing a provider LAN technology has been pursued in many forms. As presented in this article, the need to have ubiquitous simple layer 2 networks is creating a need for LAN technology.

PBB and PBB-TE, combined, provide a comprehensive native Ethernet LAN technology with an Ethernet transport capability. Each can be used alone or in combination to provide the desired service types. The standardization and deployment of PBB-TE will hopefully make this clear.

Conclusions

During the early years of bridging, Ethernet established several tenets that have been the core of its success:

- Unicast, multicast, and broadcast capabilities
- Bidirectional congruent paths for multicast and unicast traffic, and symmetry in the forward and return paths; always essential for learning, and is now so for OAM also
- Backward compatibility with legacy bridging

The solid services interface of Ethernet is desirable for providers.

The Ethernet forwarding paradigm with predictability is desirable to support provider services. Ethernet has gained many capabilities that allow Ethernet to support provider services.

The rigor that has been applied to Ethernet has allowed Ethernet to reach from the desktop to the core, albeit not as a "flat" network technology. Rather than changing the Ethernet paradigm, the capabilities are being preserved and new features added in a backward compatible fashion. Richer Ethernet topologies and engineered paths are now available within the Ethernet forwarding paradigm.

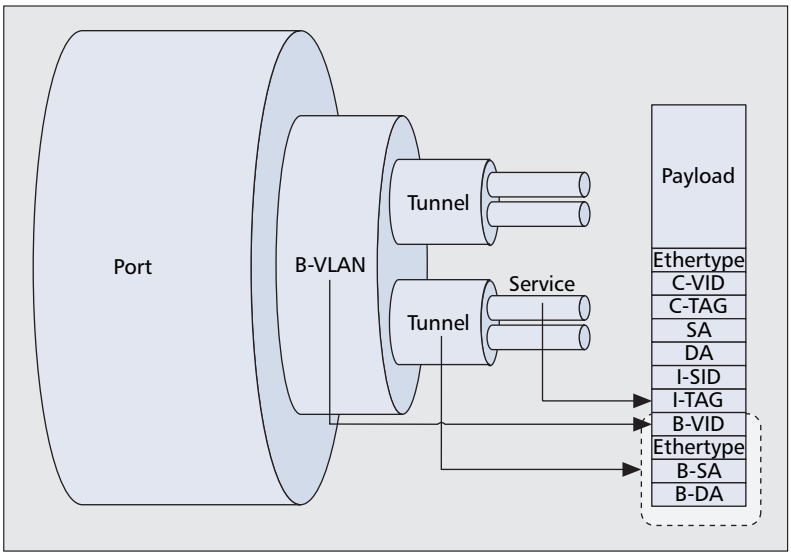

Figure 4. *Provider backbone services.*

Ethernet will continue to take an ever increasing role as part of a rich service offering. PBB and PBB-TE offer a native "Ethernet for Ethernet" infrastructure solution, combining native Ethernet without the compromises inherent in emulations, building on point-to-point technologies, and the determinism and predictability of traffic engineered point-to-point connectivity.

Acknowledgment

The authors would like to thank Paul Bottorff and Dinesh Mohan for material that inspired this article, and Nigel Bragg for a thorough review of the article.

References

[1] IEEE 802.1Q, "Virtual Bridged Local Area Networks," 2005.
[2] Paul Bottorff, Steve Haddock, Eds., "IEEE 802.1ah - Provider Backbone Bridges," Draft 4.0, Nov. 22, 2007, work in progress.
[3] IEEE 802.1ag, " Virtual Bridged Local Area Networks, Amendment 5: Connectivity Fault Management," 2007.
[4] ITU-T Draft Rec. Y.1731(ethoam), "OAM Functions and Mechanisms for Ethernet Based Networks," work in progress.
[5] P. Saltsidis, Ed., 'IEEE 802.1Qay — Provider Backbone Bridge Traffic Engineering," Draft 1.1, Dec. 22, 2007, work in progress.
[6] Don Fedyk *et al.*, "GMPLS control of Ethernet," Internet draft, May 2007, work in progress.

Biographies

Don Fedyk (dwfedyk@nortel.com) is a senior technical advisor at Nortel. He is an authority on routing system design for both connectionless and path-oriented routing. He is an active contributor to several IETF WGs in the routing area, including MPLS and CCAMP. In the past couple of years he has been focused on data and control planes for provider Ethernet. He received his B.S. and M.S. degrees in electrical engineering from the University of Waterloo, Ontario, Canada.

Dave Allan (dallan@nortel.com) has been active in data telecommunications standards for the past 12 years including WG chair roles in the DSL Forum and IETF. He has been active for over 25 years as an architect, design engineer, and developer of rea-time systems in diverse areas of technology ranging from process control and avionics to financial transaction processing. His current role at Nortel is focused on carrier infrastructure based on Ethernet and MPLS.