Ethernet as a Carrier Grade Technology: Developments and Innovations

Rafael Sánchez, Universidad Carlos III de Madrid, Spain
Lampros Raptis and Kostas Vaxevanakis, Hellas-On-Line

ABSTRACT

Recent innovations in Ethernet networking technology are enhancing both the scalability and capability of Ethernet as a carrier-grade transport technology. This article explains four main innovations recently added to Ethernet, improvements related to scalability, OAM functionality, and enhanced forwarding capability in order to permit Ethernet to assume a much larger role in carrier networks with substantial economic and operational benefits.

INTRODUCTION

Ethernet-based networking technology has become ubiquitous in both the enterprise and home broadband arenas. The combination of simplicity and rigorous specification has permitted a degree of integration and commoditization that other networking technologies have been unable to achieve.

However, some service providers' infrastructure is based on a legacy circuit-based infrastructure, using technologies like synchronous digital hierarchy (SDH), frame relay, and asynchronous transfer mode (ATM) to provide private line services and interconnection. This has placed service providers in a difficult position, as they face both the costs of supporting multiple technologies and a service arbitrage situation: they sell the same service on multiple technology platforms.

Ethernet is the technology of choice in the customer domain and is therefore a desirable choice in the service provider domain to eliminate potential interworking problems and leverage customer-driven investment. However, every technology transformation in the service provider space is time-consuming and also represents major commitment; consequently, comprehensive functionality is required as a prerequisite to mass deployment. From a carrier's perspective, Ethernet still has deficiencies with respect to operations, administration, and maintenance (OAM), reliability, traffic management, and scalability.

It turns out that many of the fundamental issues with Ethernet are well understood, and are currently being addressed with the same rigor and drive for simplicity that has been the objective of Ethernet to date. This article examines the challenges faced, including how existing Ethernet behaviors can be combined with standards in progress in order to provide a comprehensive network infrastructure that will address the carrier's concerns.

After a summary of the challenges to Ethernet, the remainder of this article is structured as follows. We describe new Ethernet technologies and how these technologies resolved some of the key challenges; we discuss traffic engineering applied to Ethernet; and finally, we cover OAM capabilities. The article concludes with the main findings, which justify the maturity of Ethernet as a carrier grade transport networking technology.

CHALLENGES TO ETHERNET

While end customers are convinced of Ethernet's cost benefits, they are demanding the same levels of performance they had from leased lines, frame relay, and ATM services. For Ethernet to reach the kind of penetration predicted by analysts, Ethernet must evolve to display the same properties as current wide area network (WAN) technologies.

The Metro Ethernet Forum (MEF) has defined this evolution as "Carrier Ethernet," which should have the following attributes.

Scalability — Providers require that the network scale to support hundreds of thousands of customers to adequately address metropolitan and regional served areas.

Protection — This really implies reliability and resiliency, as service providers typically boast "five 9s" or 99.999 percent network availability. One of the benchmark tools for achieving this has been synchronous optical network (SONET)/SDH's ability to provide 50 ms link recovery, as well as protection mechanisms for nodal and end-to-end path failures. For carrier Ethernet to be adopted — especially in support of converged real-time applications — it must match these performance levels seen by traditional WAN technologies.

Hard quality of service (QoS) — Service providers must be able to offer customers differentiated levels of service to match application requirements. QoS mechanisms provide the functionality to prioritize different traffic streams, but hard QoS ensures that service level

Lampros Raptis is currently a self-employed telecom consultant.

Kostas Vaxevanakis is currently a self-employed telecom consultant.

This article reflects the view of the authors and was finished on June 13th, 2008.

0163-6804/08/$25.00 © 2008 IEEE

Authorized licensed use limited to: BME OMIKK. Downloaded on October 28, 2008 at 13:08 from IEEE Xplore. Restrictions apply.

parameters agreed for each level of service are guaranteed and enforced across the network. This provides customers with the guaranteed deterministic performance they receive from their existing leased line services.

Service management — Service providers require mature network and service management systems that first allow quick services provisioning in order to deliver existing and new services, and second monitoring different parameters of the provided services. Such monitoring is used against an SLA, and the service provider must have the performance measurements to back up any service level claims. And if a fault does occur, the service provider needs to have the troubleshooting functionality to locate the fault, identify which services have been impacted, and react appropriately.

Time-division multiplex (TDM) support — While service providers see substantial growth potential in Ethernet services, existing leased lines are still a significant revenue source for them that they must be able to retain and seamlessly interwork with existing leased lines services as they migrate to a carrier Ethernet network.

Equipment vendors are challenged with how to add this carrier-grade functionality to Ethernet equipment without losing the cost effectiveness and simplicity that make it attractive in the first place. In the next sections we examine the different technologies designed to achieve this.

Ethernet Technologies

The MEF has defined Ethernet services [7] using the concept of Ethernet virtual connections (EVCs) established across an Ethernet network. Customer equipment (CE) attaches to the network at the user-network interface (UNI) using standard 10 Mb/s, 100 Mb/s, 1 Gb/s, or 10 Gb/s Ethernet interfaces. There are three types of EVCs defined:

- Point-to-point, called E-Line
- Point-to-multipoint, called E-Tree
- Multipoint-to-multipoint, called E-LAN

In order to provide such services, different Ethernet technologies have been proposed and are used for the delivery of the previous services.

IEEE 802.1Q Virtual LAN

The basic technology standard used for delivering an E-LAN service is the IEEE 802.1Q standard [1] for virtual LANs (VLANs). This standard creates VLANs across a common LAN infrastructure to enable enterprises to support and separate traffic from different departments within a company (e.g., finance, legal, and general administration). Each VLAN is identified by a Q-tag (also known as a VLAN tag or VLAN ID) that identifies a logical partitioning of the network to serve the different communities of interest.

IEEE 802.1Q works fine within the boundaries of a single organization, but is found to be inadequate when service providers attempt to deliver Ethernet services to multiple end users over a shared network infrastructure. Issues arise because enterprises need to retain control over their own VLAN administration (e.g., assigning Q-tags to VLANs), and over a shared infrastructure the service provider must control this to ensure that one customer's Q-tags do not overlap with another's. Also, because the Q-tag consists of a 12-bit tag, up to 4094 possible service instances can be created. (Note: 4096 service IDs are available, but two of these are reserved for administration.) Although this is sufficient for an enterprise's LANs, it does not offer the scalability required to support Ethernet services in a large metropolitan area. What is needed is a method for defining secure Ethernet services to individual customers within which each customer can create further LANs for departments or groups of users. There are two developing standards that support this approach: IEEE 802.1ad provider bridges [2] (also known as Q-in-Q or VLAN stacking) and IEEE 802.1ah provider backbone bridges [3] (also known as MAC-in-MAC).

The standardization of these technologies is being driven by the IEEE 802.1 working group. The provider bridges standard was officially approved in December 2005, the provider backbone bridges standard in June 2008.

The standardization of these technologies is being driven by the IEEE 802.1 working group. The Provider Bridges standard was officially approved in December 2005, while Provider Backbone Bridges standard was officially approved in June 2008.

IEEE 802.1ad Provider Bridges

Provider bridges work by simply adding an additional service provider VLAN ID (S-Tag) to the customer's Ethernet frame. This new S-Tag is used to identify the service in the provider network, while the customer's VLAN ID (C-Tag) remains intact and is not altered by the service provider anywhere within the provider's network, as shown in Fig. 1. This allows the C-Tag to be transparent within a Q-in-Q network.

Provider bridges use the S-Tag to identify the service to which a customer's Ethernet frames belong; therefore, each service instance requires a separate S-Tag. Because the S-Tag consists of a 12-bit tag, provider bridges have the same scalability limitation as IEEE 802.1Q: only 4094 services instances can be created.

In addition, provider bridges uses the same medium access control (MAC) address for the provider's and customers' networks. This makes both networks appear as one large network to the provider's switches, as shown in Fig. 2.

In the scenario depicted in Fig. 2, the provider's and customers' MAC addresses are visible to all network elements of the service provider; this creates a significant burden for core switches, as they must maintain a forwarding table for every MAC address in the service provider and customer networks. Also, any changes to the customer network will have an impact on the provider core. For example, when a new host is added in the customer's network, the new MAC address must be learned by the provider's switches, or when a failure occurs in the customer network, the resulting action taken by Spanning Tree Protocol (STP) can impact the provider network. Although such changes are outside the service provider's network, they impact their network and create a temporal situation until spanning tree convergence is achieved. From the customers' perspective, a potential security concern emerges from the fact that their addressing information is visible outside of their secure network domain, since there is no separation between customer and provider MAC addresses.

Authorized licensed use limited to: BME OMIKK. Downloaded on October 28, 2008 at 13:08 from IEEE Xplore. Restrictions apply.

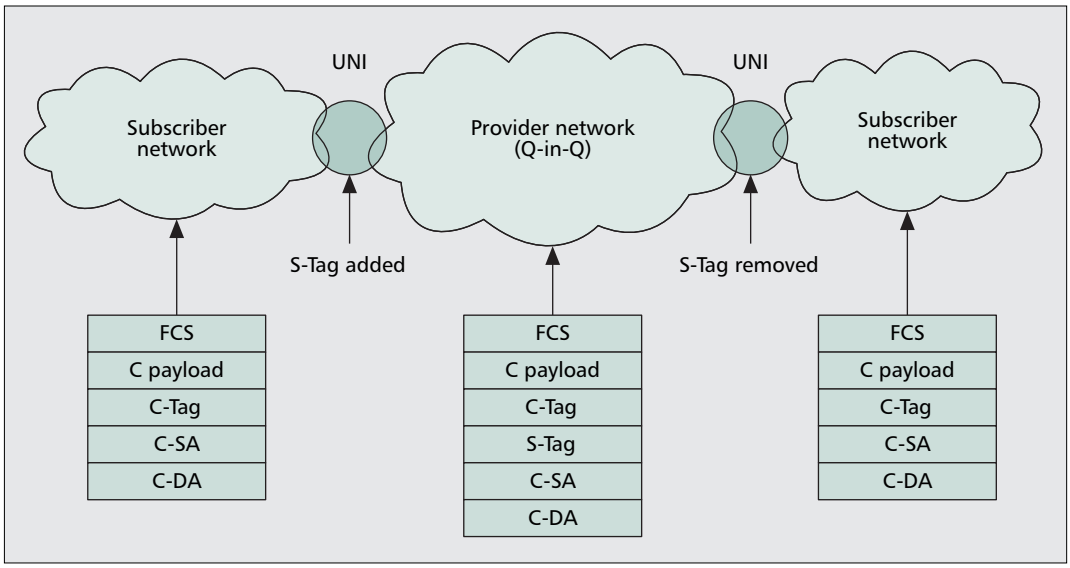


Figure 1. *S-Tag added to the customer frame.*

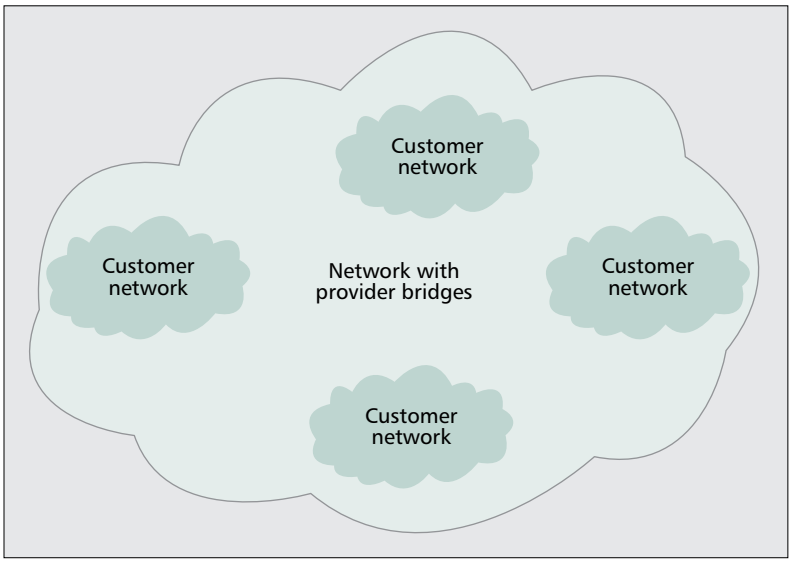


Figure 2. *Provider's and customers' MAC addresses visible to all networks.*

Provider bridges do not provide separation between the provider and customer networks, which creates problems where control protocols are concerned. Most Ethernet control protocols, such as those transporting bridged protocol data units (B-PDUs) used by customer networks, could not interact with the provider's networking equipment in traditional Ethernet. B-PDUs were identified by their destination MAC address and do not have a VLAN tag associated with them. For example, STP was identified by destination MAC address 01-80-C2-00-00-00. Provider bridges could not provide differentiation between customer and provider B-PDUs because each entity's B-PDUs used to have the same MAC address, and duplicate MAC addresses could not be supported. This caused unpredictable network behavior because the provider's networking equipment cannot distinguish between customer and provider B-PDUs. The IEEE standard solves this limitation by introducing a different set of destination MAC addresses for B-PDUs in the provider's network (01-80-C2-00-00-08). However, to support these new provider B-PDU MAC addresses, the service providers must upgrade (in some cases even replace) the existing Ethernet switches, because B-PDU MAC addresses are not configurable. For this reason, provider bridge technology has significant limitations for E-LAN services that must support multiple customer control protocols.

IEEE 802.1ah Provider Backbone Bridges

Provider backbone bridges (PBBs) (IEEE 802.1ah) evolves the Ethernet frame by adding a MAC header dedicated to the service provider and, in doing so, adds a backbone source and destination MAC address, a backbone VLAN ID (B-Tag), and a backbone service ID (I-Tag) to the customer's Ethernet frame. Figure 3 illustrates the PBB frame and shows how this compares to the standard Ethernet frame (IEEE 802.1), VLANs (IEEE 802.1Q), and provider bridges (IEEE 802.1ad).

One of the main benefits of PBBs is that *the 24-bit I-SID (I-Tag) identifies the service in the PBB network (PBBN)*. This means PBBs provide *up to 16 million services*, completely removing the scalability problems of provider bridges (or Q-in-Q).

In addition, PBB provides clear separation between the service provider (Q-in-Q) and the PBBN, because each has a dedicated set of MAC addresses, as shown in Fig. 4. When an Ethernet frame reaches the Ethernet UNI, the service provider MAC address is added to the incoming Ethernet frame, and PBBN switches check this MAC address against their forwarding tables. This is an added advantage in that only switches at the edge of the PBBN need to be PBB-enabled. Switches in the core of the PBBN are just provider bridges. In this case the incoming S-Tag has the same Ethertype as the PBB B-Tag, and forwarding is based on the B-DA and B-Tag.

This solution allows MAC addresses from the

Authorized licensed use limited to: BME OMIKK. Downloaded on October 28, 2008 at 13:08 from IEEE Xplore. Restrictions apply.

PB networks (Q-in-Q) to overlap with the PBB MAC addresses, because the incoming frames are tunneled by PBBs and are not used when switching frames inside the PBB network. As a result, PB devices are free to assign identifier and class of service values to their VLANs (C-VID and S-VID) without any concern that those VLANs will be altered by the PBBN. Meanwhile, the PBBN does not need to worry about coordinating VLAN (S-VID and C-VID) administration with its customers (in this case the provider bridge networks).

Also, because the PBB core switches only use the PBB MAC header, there is no need for them to maintain visibility of provider bridges' MAC addresses, reducing the burden on the forwarding tables in the PBB network. This also ensures that changes to PB networks have almost no impact on the PBB network (except in the edge), improving the stability of the PBB network. Finally, security is improved, because the PBB switches in the core are no longer inspecting the PB MAC headers.

Another benefit of PBBs is that because *the I-SID is used for service identification, the B-Tag can be used to segregate the PBB network into virtual networks* (or regions) with different technologies and capabilities on each. For example, some VLAN ID can be assigned to provide E-LINE services with traffic engineering, while other VLAN-IDs provide E-LAN/E-TREE services with no traffic engineering. Backbone VLANs enable the support of multiple services instances; for example, a B-VID can be engineered to support 1000 10 Mb/s E-Line services between points of presence (POPs), as in Fig. 5.

This means the service provider engineers the PBB network once when the B-VID is set up. Individual services can then be activated at the source and destination nodes, and supported over the B-VID according to its engineered limitations. With provider bridges, each individual service needs to be configured across the network node by node, creating a substantial operational burden.

Since PBBs tunnel PB frames, all Ethernet control protocols (B-PDUs) are tunneled transparently across the PBB network. This allows Ethernet control protocols to be used independently by the PB and PBB networks.

Adding Traffic Engineering to Ethernet

It is now possible to support traffic engineering using native Ethernet with a new technology called PBB — traffic engineering (PBB-TE). PBB-TE is an innovative Ethernet technology, originally introduced to the IEEE by Nortel and BT, and now an IEEE project, PBB-TE P802.1Qay [6]. This technology proposes only minor addition to the existing Ethernet standards. In its simplest form, PBB-TE provides Ethernet tunnels that enable deterministic service delivery with the TE, QoS, resiliency, and OAM requirements service providers demand.

PBB-TE takes advantage of the fact that by simply turning off some Ethernet functionality, the existing Ethernet hardware is capable of a new forwarding behavior. This means that *a connection-oriented forwarding mode can be introduced to current Ethernet networks* without complex and expensive network technologies.

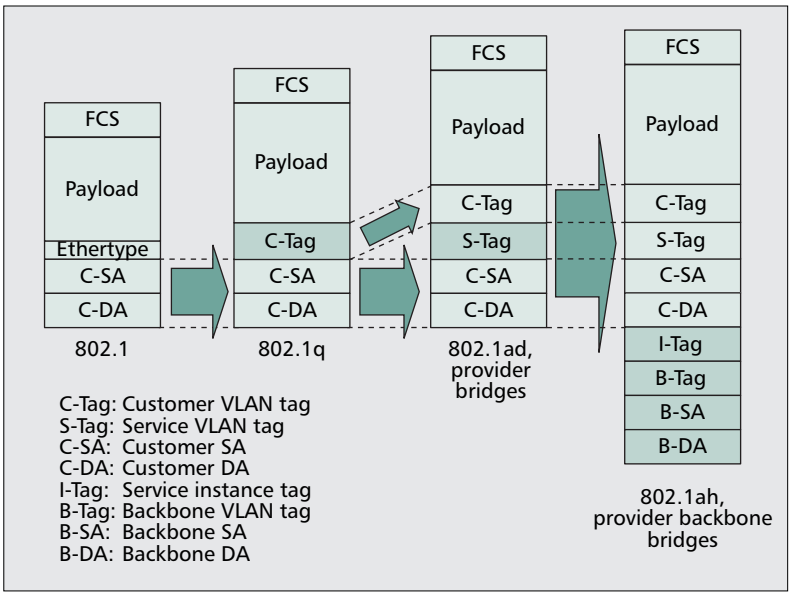


Figure 3. *PBB Ethernet frame.*

Currently, Ethernet switches forward on the basis of a full 60-bit lookup of both the VLAN tag (12 bits) and the destination MAC address (48 bits) in each Ethernet frame. In conventional operation both the VLAN ID (VID) and MAC address are globally unique, but this does not have to be the case. Where a VID typically identifies a loop-free multicast domain in which MAC addresses can be flooded, if we choose to configure loop-free MAC paths instead, the VID is freed up to be used to denote something else. In the case of PBB-TE it will use a set of VIDs to identify specific paths through the network to a given destination MAC address. Each VID is then locally significant to the destination MAC address only, and since the MAC address is still globally significant, the combination of VID + MAC (60 bits) becomes globally unique.

PBB-TE allocates a range of VID/MAC addresses whose forwarding tables are populated via the management or control plane instead of through the traditional flooding and learning techniques. In this case STP and all its associated constraints and problems disappear. The switches still behave fundamentally as with traditional Ethernet: forwarding data to its intended destination. What is different is the fact that the forwarding information is no longer based on the MAC learning mechanisms of the switches, but is provided directly by the management plane, resulting in a prescribed predetermined path through the network and totally predictable network behavior under all circumstances.

In the example shown in Fig. 6, two unidirectional paths have been configured between provider edges (PEs) 1 and 2 (a pair of links in opposite directions is required for bidirectional connectivity). Each PE is IEEE 802.1ah enabled, allowing the service provider to clearly separate the service provider and customer MAC domains,

Authorized licensed use limited to: BME OMIKK. Downloaded on October 28, 2008 at 13:08 from IEEE Xplore. Restrictions apply.

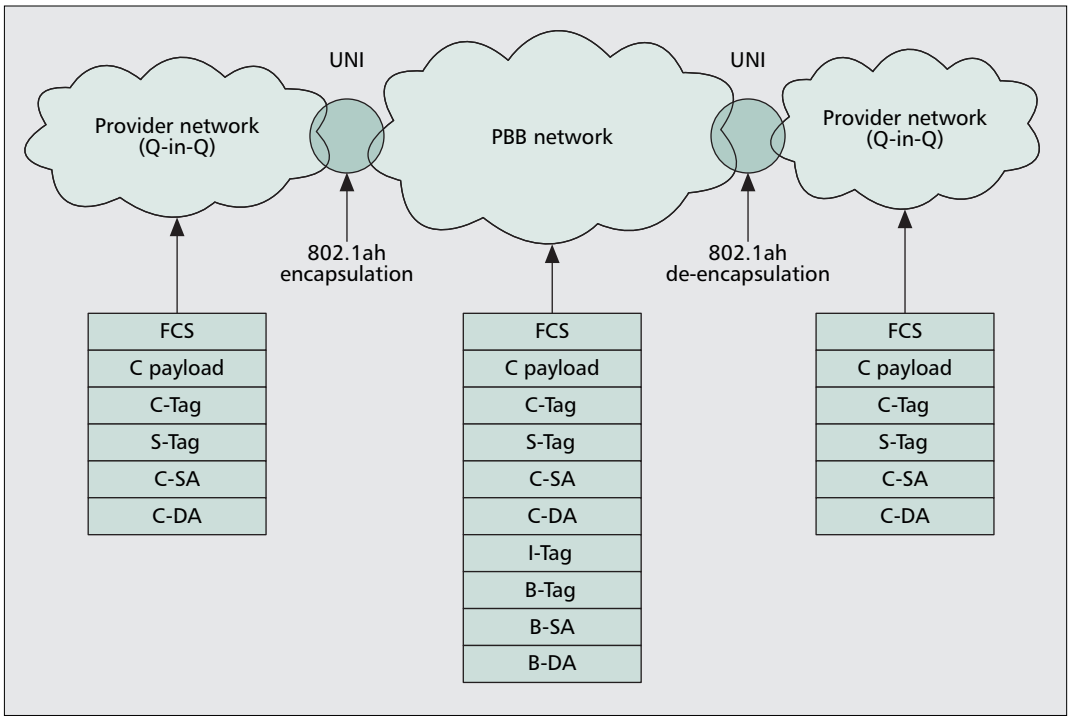


Figure 4. *MAC addresses separated at the UNI.*

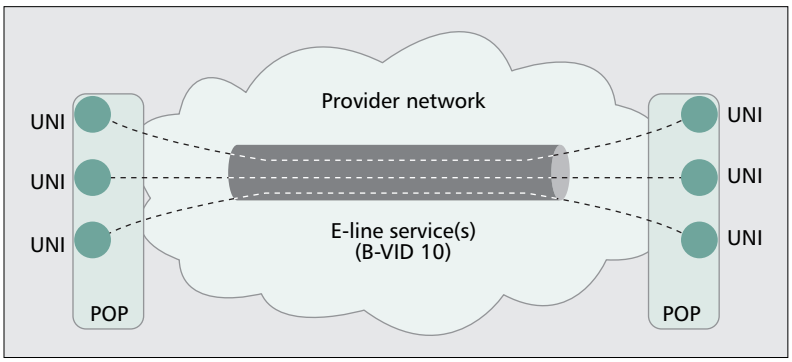


Figure 5. *Single B-VLAN for multiple services.*

thus allowing the service provider to apply PBB-TE within the core of the network. Within the service provider domain, a number of VIDs have been reserved for PBB-TE; these include VIDs 44 and 45 in our example. As explained, within the group of VIDs reserved for PBB-TE behavior, the VID is used, together with the MAC, to identify common paths toward the same destination. Instead, VIDs 44 and 45 are used to separately identify the two paths between PEs 1 and 2. Both of these VIDs can be reused to create paths between a different pair of PEs because it is the combination of MAC and VID that uniquely identifies each of these paths.

PBB-TE preserves the destination-based forwarding attributes of Ethernet, which means multiple sources can use a VID+MAC destination. If 16 VIDs were reserved for PBB-TE in this network, the network could be fully meshed 16 times. This would provide massive scalability for the PBB-TE links and still leave 4078 VIDs for normal connectionless Ethernet behavior operating on the same network. It should be noted that each frame still carries a source MAC address that uniquely identifies its origin; so PBB-TE offers the scaling of destination-based forwarding in the core (order *N*) while preserving the operational attributes of point-to-point at the edges.

In the example in Fig. 6, a pair of bidirectional Ethernet links has been configured across the network to create working and protection paths (they would typically be diverse routed; however, in our example, they were made to cross in a core switch to shown how different VIDs may be used to identify different routes). PBB-TE derives connection monitoring from IEEE 802.1ag (connectivity fault management) messages. A connectivity check (CC) session is established on both paths. Both ends of the link send CC frames at regular (configurable) 10 ms intervals and listen to the messages that arrive. If three CC messages do not arrive, the link is deemed to be down, and a protection switch is initiated. Alternatively, alarm indication signal (AIS) messages defined by ITU-T Y.1731 could be used to trigger a protection switch.

Protection switching [6] is implemented by applying the new VLAN tag (that of the protection path) to each frame at the encapsulation point. The control plane is used to configure and monitor the paths, but is not involved in the actual switching, so sub-50-ms protection switching (similar to SONET/SDH) can be achieved.

Adding OAM to Ethernet

OAM functionality in traditional TDM networks is well defined and an important building block in ensuring that operators can deliver "carrier grade" performance services. The traditional Ethernet in the LAN environment does not have the OAM functionality required by network operators in MAN and WAN environments.

Authorized licensed use limited to: BME OMIKK. Downloaded on October 28, 2008 at 13:08 from IEEE Xplore. Restrictions apply.

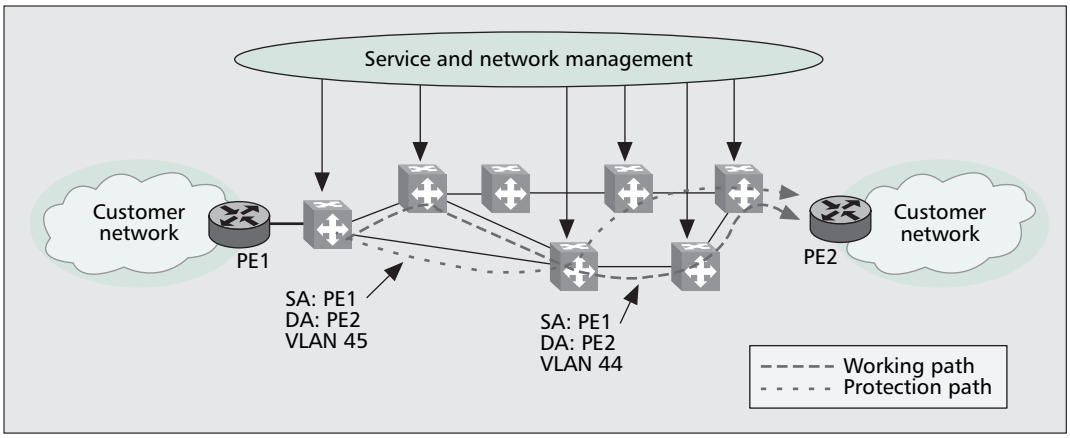


Figure 6. *PBB-TE configuration.*

If carrier Ethernet is to fulfill its promise as the next-generation packet-based infrastructure for MANs and WANs, OAM capabilities must be added to Ethernet.

New standards that provide Ethernet with OAM capabilities are described next.

Fault Management

There are two main areas of OAM: fault management and performance monitoring. Fault management ensures that when a defect occurs in the network, it is reported to the operator, who can then take the appropriate action. This is divided into the following functions:

1 **Fault detection** — IEEE 802.1ag [5]and ITU-T Y.1731 [4] support fault detection through continuity check messages (CCMs). These allow endpoints to detect an interruption in service. CCMs are sent from the source node to destination node at periodic intervals; if either end does not receive a CCM within a specified duration, a fault is detected against the service.

2 **Fault verification** — IEEE 802.1ag and ITU-T Y.1731 support fault verification through loopback messages (LBMs) and loopback reply (LBR). These can be used during initial setup or after a fault has been detected to verify that the fault has occurred between two endpoints.

3 **Fault isolation** — IEEE 802.1ag and ITU-T Y.1731 support fault isolation through linktrace messages (LTMs) and linktrace reply (LTR). In our example (Fig. 7) node A initiates an LTM; each intermediate node along the path to F (B and E) sends an LTR back and forwards the LTM toward node F. Under normal conditions, it allows the operator to determine the path used by the service through the network, whereas under fault conditions, it allows the operator to isolate the fault location without making a site visit.

4 **Fault notification** — ITU-T Y.1731 supports fault notification through the AIS. In our example (Fig. 8), a failure between nodes B and E triggers AIS packets in both directions toward the service endpoints. This functionality alerts the operator of a fault in the network before it is reported by customers. At nodes A and F, the service endpoints, the alarm can be replicated across all services supported at that UNI impacted by the fault. The AIS packets are issued periodically by nodes B and E to ensure that while the fault still exists, a failure state is maintained. Additionally, the AIS packets can be used to trigger the survivability mechanisms.

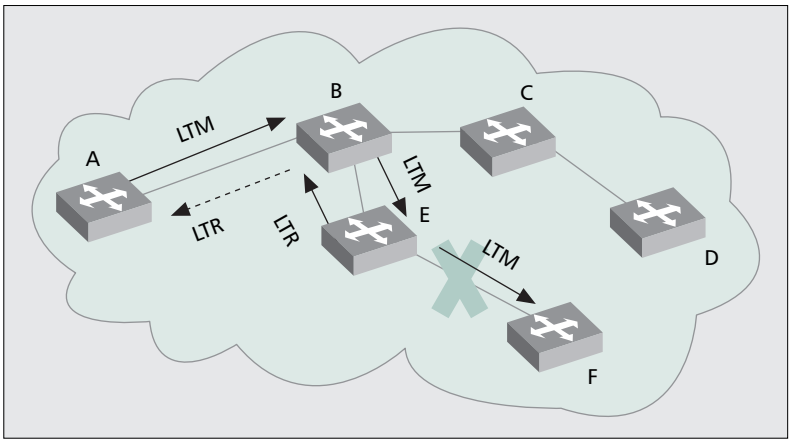


Figure 7. *Fault isolation.*

Performance Monitoring

In many respects the fault management concepts above have been adopted from existing practices in traditional TDM networks. However, while connection-orientated TDM services offer customers predictable and guaranteed service, packet- or frame-based services are usually connectionless and can have varying performance levels. This is because each individual frame in a service can suffer varying delays due to possible queuing, while network congestion can result in actual loss of frames. Video and voice services, which are part of a residential triple play bouquet, are particularly susceptible to the effects of latency and jitter. As a result, carrier Ethernet networks require advanced performance monitoring to help service providers to determine that SLAs are met; this functionality is introduced by ITU-T Y.1731. The following functionality is included:

1. **Frame loss ratio** — ITU-T Y.1731 calculates frame loss by sending transmit and receive counters within the CCM for dual-ended measurements. The far end counters can then be compared to those produced locally to derive frame loss as a percentage.

Authorized licensed use limited to: BME OMIKK. Downloaded on October 28, 2008 at 13:08 from IEEE Xplore. Restrictions apply.

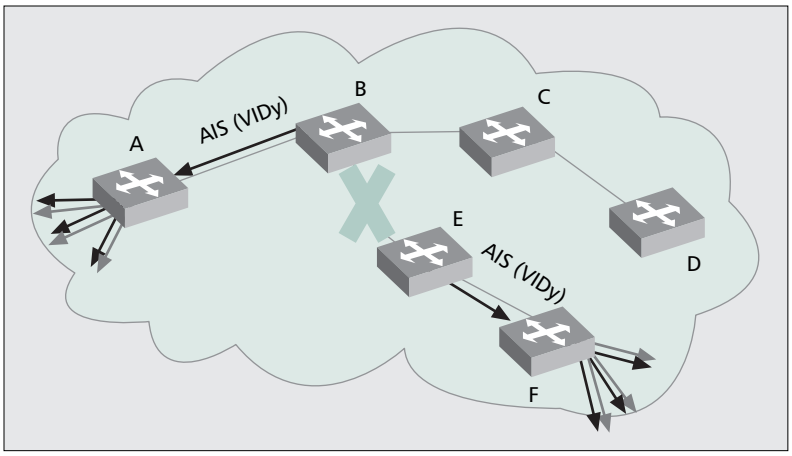


Figure 8. *Fault notification.*

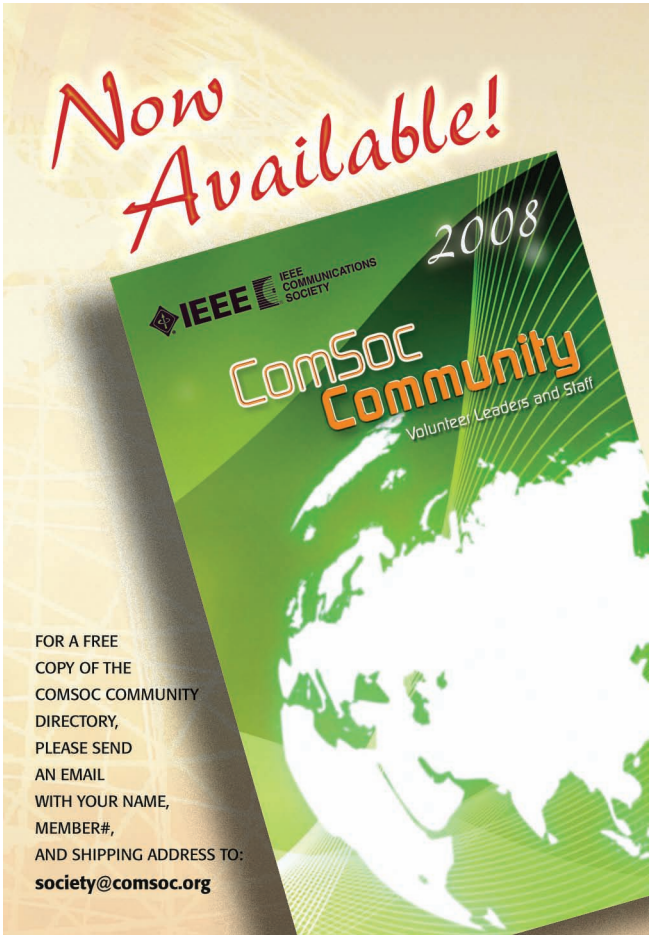
2. **Frame delay** — Similarly, ITU-T Y.1731 calculates frame delay (or latency). The receiving end can derive the time delay experienced across the network. This requires each service endpoint to have synchronized clocks.
3. **Frame delay variation** — Finally, ITU-T Y.1731 calculates frame delay variation (or jitter) by tracking frame delay measurements.

The emergence of carrier-grade Ethernet has driven the need for improved Ethernet OAM functionality. This functionality is normally collected by management systems. Without this capability, it is impossible to provide the comprehensive network management functionality operators have today in their TDM networks.

Conclusions

Traditionally, Ethernet lacks some capabilities to become a technology deployed in the metropolitan and wide area network environments. However, recent innovations like PBB, PBB-TE, and OAM allow operators to consider Ethernet as a carrier grade networking technology alternative to traditional technologies like SONET/SDH, ATM, and MPLS.

Provider backbone bridges (IEEE 802.1ah) provide carrier-grade scalability and improved security due to the separation between service providers' and customers' addressing schemes. Provider backbone bridging — traffic engineering is then employed in the service provider domain, creating the ability to configure resilient SLA-driven point-to-point Ethernet trunks. Finally, the combination of IEEE 802.1ag [5] and ITU-T Y.1731 [4] provides powerful fault management and performance monitoring capabilities to Ethernet.



References

[1] IEEE . 802.1Q, "Local and Metropolitan Area Networks, Virtual Bridged Local Area Networks," 2003.
[2] IEEE 802.1ad, "IEEE Draft Standard for Local and Metropolitan Area Networks, Virtual Bridged Local Area Networks, Amendment 4: Provider Bridges."
[3] IEEE 802.1ah, "IEEE Draft Standard for Local and Metropolitan Area Networks, Virtual Bridged Local Area Networks, Amendment 6: Provider Backbone Bridges."
[4] ITU-T Rec. Y.1731, "OAM Functions and Mechanisms for Ethernet based Networks," 2006.
[5] IEEE 802.1ag, "Local and Metropolitan Area Networks, Virtual Bridged Local Area Networks, Amendment 5: Connectivity Fault Management."
[6] IEEE P802.1Qay, "Provider Backbone Bridge Traffic Engineering."
[7] MEF 6: Ethernet Services Definitions — Phase I.

Biographies

Author biographies were not available at the time this issue went to press.

Authorized licensed use limited to: BME OMIKK. Downloaded on October 28, 2008 at 13:08 from IEEE Xplore. Restrictions apply.