Carrier Ethernet

Andy Reid, Peter Willis, Ian Hawkins, and Chris Bilton, BT

ABSTRACT

This article sets out the opportunity and the role of Ethernet within large-scale carrier networks. Carrier Ethernet can greatly reduce the consequences of the complexity associated with the large scale and broad scope of carriers' networks by being a cost-effective replacement for SONET/SDH. However, in order to achieve this, carrier Ethernet needs to provide the equivalent level of transparency, simplicity, and reliability currently achieved by SONET/SDH, and the emerging IEEE Ethernet standards for PBB and PBB-TE are well suited to this role. These technologies are an ideal complement to IP/MPLS for which they can provide highly cost-effective managed and guaranteed bit pipes.

BACKGROUND

The introduction of Ethernet technology by carriers encounters several challenges. There are the classic challenges of making a positive business case and developing a realizable overall network evolution plan. However, the introduction of Ethernet technology introduces an extra challenge since Ethernet, being a packet technology, has significant differences from the circuit-switched synchronous optical network/digital hierarchy (SONET/SDH) technology it would most naturally replace. The challenge is to develop the Ethernet technology in such as way that it can be a practical functional alternative for carriers for their existing SONET/SDH networks. This article looks at the way Ethernet technology can meet this challenge.

A key benefit of Ethernet as a technology is that it meets the needs of a wide mass market and so has a high volume production of basic hardware and equipment. To use this technology, however, the carrier is left with the dilemma of choosing between:

- A volume — and lower-cost — technology that is not ideally suited to carrier requirements
- A carrier-specific technology that misses the cost advantages associated with high volumes (Fig. 1).

In fact, two technologies currently dominate the high-volume enterprise and consumer markets: IP routing and Ethernet bridging. Each of these has found its place, even if there is some overlap: very broadly speaking, IP routing has resulted in volume software code, while Ethernet bridging has led to volume switching hardware.

For many years now, and again broadly speaking, the carrier industry has been exploiting the enterprise IP software base, but running it on special high-speed hardware and with some additional functionality, mainly under the banner of multiprotocol label switching (MPLS). In more recent years, the carrier industry has turned its attention to the exploitation of the volume Ethernet switching hardware within the carrier networks.

THE PLACE FOR EXPLOITING ETHERNET IN CARRIERS' NETWORKS

IP, with the help of MPLS extensions, provides the basis for a wide variety of carrier services. For example, as well as supporting basic Internet services, IP/MPLS provides for virtual private networks, traffic engineering, and the support of virtual private circuits (also called pseudowires). However, many carriers are finding that while MPLS offers many features for the carrier, enabling them all at the same time on the same network introduces a whole new set of complex challenges.

For example, IP/MPLS technology has the ability to offer a wide variety of services from a single network platform and thus would appear to give both economies of scale — the bigger it is, the cheaper per unit — and economies of scope — the more services covered by the platform, the cheaper it becomes per platform. However, this does not continue forever, and after a certain size the growing complexity of managing the large scale and/or broad scope starts to show diseconomies of scale and scope.

Diseconomies of scale and scope are normally closely coupled to the complexity of the network and its operations. In this context, complexity is usefully defined in terms of the scaling of interactions between components of a system as more components are added to the system. Complexity arises when the number of interactions grows more than proportionately with the number of components. Without a determined overall structure to the system, this is very likely, as in a system where components generally interact with each other, the number of interactions grow as the square of the number of components [1].[1]

[1] *Economists refer to the "cost of communication" between the people or elements of a large system as being a significant source of diseconomies of scale and scope. This is described in many texts on microeconomics.*

0163-6804/08/$25.00 © 2008 IEEE

So when either new functional components are added to a system like MPLS, and/or more equipment is added to a deployed network, the number of interactions between all the components is likely to grow more than proportionately. There are many pertinent examples, many of which will fall into the following categories:

- Scope: The increasing functional scope and "feature creep" of systems can lead to increasingly costly testing of feature interactions.
- Scale: The interactions between the individual components of a distributed system can give rise to rising costs and technical limits (e.g., the amount of equipment in a single routing area).
- Within life cycle: The development, deployment, and operations of large complex systems can have interactions across this life cycle that can generate costs which grow disproportionately with the scale and scope of the system.
- Between life cycles: When one system development needs to interact with another system that has an independent life cycle, it can generate further disproportionate costs.

All of these are likely to give rise to unavoidable increases in cost that are disproportionate to the size of the system, many of which may have been unforeseen at the outset.

Practical experience suggests that the interactions across the life cycle are particularly significant. Unforeseen interactions, or interactions that were insufficiently specified at an early stage in the life cycle cause considerable delay, additional cost, and curtailment of features at later stages in the life cycle. This is particularly true for telecommunications operators as there is a general expectation of continuous service. That is, operators and their customers usually require that any new technology must be integrated into an existing network without interrupting existing services.

When the diseconomies of scale are taken into account (Fig. 2), there is actually an optimal size and scope for a network.

The primary way of avoiding this growing cost of complexity is to "divide and conquer." That is, by partitioning out specific areas of functionality in such a way as to minimize the interdependence between these separated areas, the cost of complexity can be greatly reduced. Central to this strategy is the lack of interdependencies between the separated areas; that is, in order to "conquer," the "divide" needs to be real rather than cosmetic.

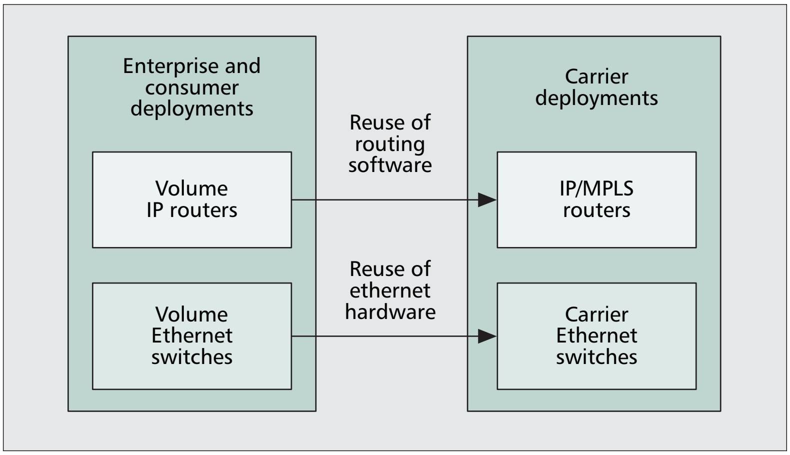


Figure 1. *Reuse of high-volume technology by carriers.*

The scaling of costs is illustrated in Fig. 2. A single "monolithic" design can minimize the initial fixed costs of a network as there is minimal duplication. However, it is most likely to encounter early diseconomies of scale and scope as the network grows. Although a partitioned approach may well have a higher initial fixed cost as there may be duplication of costs across each area, it is likely to better handle increasing scale and scope.

Carrier networks have taken full advantage of this in the past by, in particular, separating switching from transmission.[2] Switching has focused on service-oriented features using signaling systems whilst transmission has concentrated on the cost effective management of bandwidth on the assumption that the managed capacity is largely static with the configuration of the transmission paths lasting months even years.

A great advantage to the transmission area is that it has made no assumption on the higher-layer switching technology, thus minimizing any dependence between any transmission network technology and any switching network technology. This is borne out by the fact that the current transmission technologies, including SONET/SDH and optical transport network (OTN),[3] have supported the full range of switching technologies: public switched telephone network/integrated services digital network (PSTN/ISDN), voice and data mobile, third-generation (3G) mobile, as well as various data network technologies including asynchronous transfer mode (ATM), frame relay, IP, MPLS, and most lately Ethernet bridging, without any inherent change to the transmission technology. Moreover, all of these switching technologies can be supported in parallel on the same transmission network at the same time. This is illustrated in Fig. 3.

[2] *The choice of terminology here should be taken in the context of this article. This split arises out of countering diseconomies of scale, and reflects organizational, operational, as well as technical splits. For many carriers, historically, switching and transmission are the names that have been given to these organizational structures. Several terms could be appropriate; notably ,transmission as used in this article is often referred to as* transport*. Transport, however, has precise definitions in some standards bodies, but unfortunately not all the same. (For example. the ITU-T SG15 definition is quite different from the ITU-T SG13 definition used in NGNs to separate service and transport, while OSI layer 4 transport is quite different again). As used in this article, switching, in quotes, can be loosely associated with routing and NGN style control planes, while transmission, in quotes, can be loosely associated with ITU-T SG15's transport.*

[3] *OTN is a suite of standards allowing networking with cross-connects of basic signals carried by optical wavelengths multiplexed with dense wavelength-division multiplexing (DWDM).*

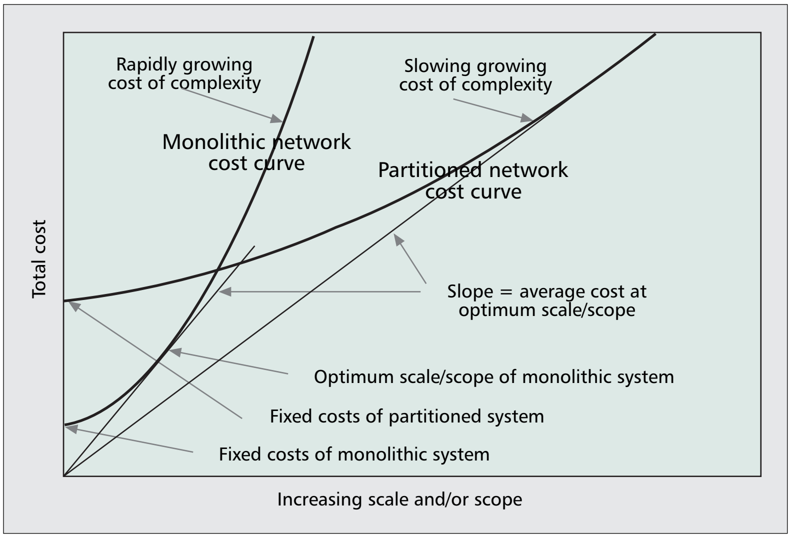


Figure 2. *Economies and diseconomies of scale and scope.*

The creation of independence in this way is often referred to as client/server network layering. This is a foundation of the functional architecture for SONET/SDH [2]. While this originally dealt with circuit switched networks, ITU-T has extended this functional architecture approach to packet networks [3].

This split between switching and transmission technologies has given a very cost effective solution for managing carrier networks. In many cases, the main additional cost of this architecture has been the cost of the separate network management systems optimized to the different demands of switching and transmission networks. However, compared to the ability to develop and deploy new switching technology, or upscale existing switching technologies without introducing new transmission, this OSS cost has been recovered many times over.

Another flexibility arising out of the split between switching and transmission is the range of business models it supports. Many organizations wish to deploy their own switching network independent of a carrier's switching network. Such organizations include corporations, public service and government agencies, as well as other network service providers. By purchasing transmission services from the carrier, these organizations can determine their own switching technology, its features, and its scale and scope. In many countries now, this split is also the basis of practical regulation to promote and sustain competition.

Ethernet in the enterprise space is oriented towards simple but highly cost effective and reliable switching of traffic. This suggests that if Ethernet technology can be reused to meet the carrier's transmission network requirements, it should have a role in replacing the SONET/SDH crossconnect technology in the next generation of carriers' transmission networks.

Distinguishing Ethernet Services from Ethernet Technology

The flexibility created by a transmission network means that a carrier can use the transmission network to both manage the internal complexity as well as offer transmission services to other organizations running their own networks. The presentation of these transmission services can vary according to the needs of the customer's technology and protocols. This means that the technology and protocols of the service interface may well not be the same as the technology which is used to build the transmission network. SONET/SDH and OTN are current technologies for building transmission networks, however, they support many types of non transmission service interfaces including Frame Relay, ATM, as well as Ethernet.

Therefore one of the objectives in reusing Ethernet technology for building a carrier transmission network is that it must support a reasonably wide variety of types of customer interface.

Conversely, Ethernet services — that is the transport of customer Ethernet traffic — are widely supported on a variety of network technologies including ATM, SONET/SDH, and OTN.

Ironically, the one type of service interface that has, until recently, been difficult to support with a transmission network built with Ethernet technology are Ethernet services. The reason is

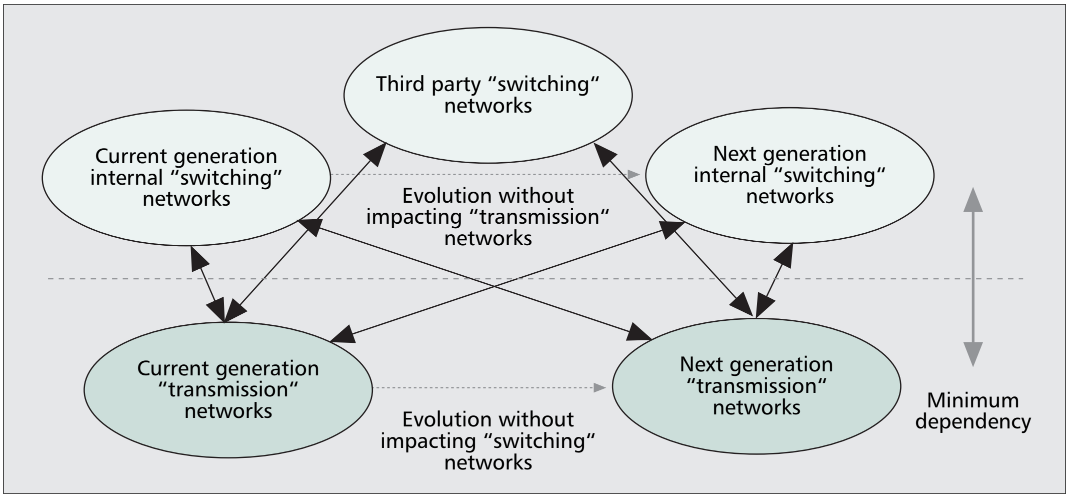
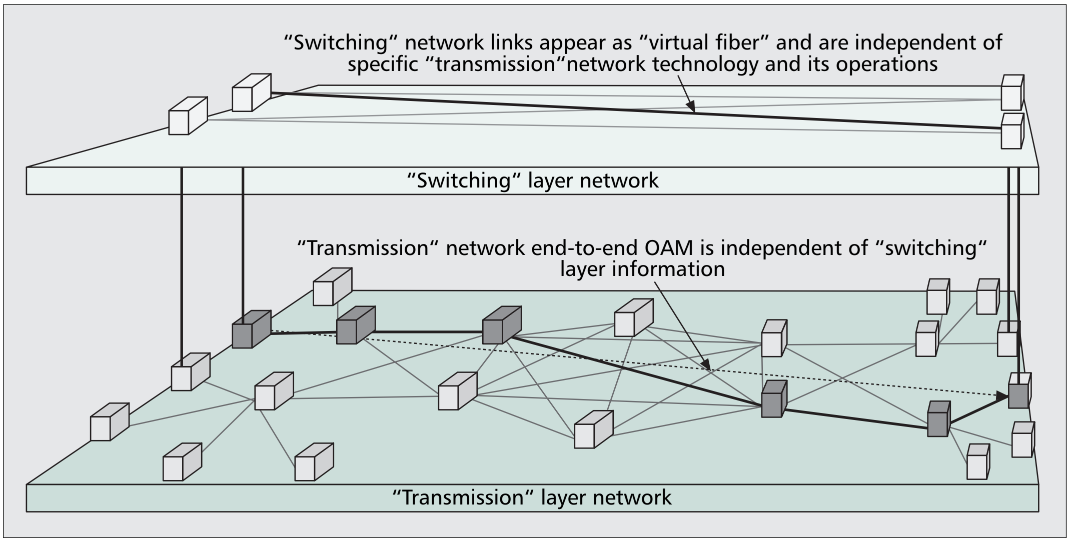


Figure 3. *Network evolution enabled by "transmission" networks.*



■ **Figure 4.** *Data transparency.*

The principle of data transparency when coupled with the need for any transmission network to manage its own reliability, means that the transmission network must only support point to point or simple, statically replicated point to multipoint services.

straight forward; Ethernet has one frame header and any fields in the header used by the customer cannot be used by the carrier and vice versa. The recently completed IEEE 802.1ah standard for provider backbone bridging (PBB) [4] overcomes this difficultly by allowing the carrier to add a completely new Ethernet medium access control (MAC) header and carry the customer's MAC Ethernet header transparently.

TRANSPARENCY, THE KEY TO INDEPENDENCE

The main objective of a transmission network is to achieve economies of scale and scope in a carrier's network without introducing the diseconomies of scale and scope associated with complexity. And the key feature in achieving this is the independence of any transmission network from any higher-layer switching network. There are a number of perspectives on independence, including both functional independence and operational independence. However, two primary aspects of this independence that are worthy of closer consideration are data transparency and quality of service (QoS) transparency. They are mainly concerned with functional independence, however, operational independence is also important.

DATA TRANSPARENCY

In order to be independent of any switching network, a transmission network must not make any assumptions about what data the switching network is sending. If the switching network in the course of time decides to change the meaning of any data, the transmission network must be unaffected by this. Moreover, if it becomes important to carry some new switching technology, this is only possible if the transmission network makes no decisions based on any of the switching network's data. In the language of many engineers, the transmission network must never "snoop" the payload. This is the principle of data transparency.

This principle of data transparency has some subtle but profound consequences. In order to be independent, the transmission network must have independent connectivity management/control and must not rely on any switching network information (e.g., addresses). In practice, this means that the destination (or set of destinations in the case of point-to-multipoint[4]) of data must be implicitly specified by the port through which it passes into the transmission network. Equally, at the destination, traffic can arrive from a number of sources, but identifying the originating source must only rely on information generated within the transmission network.

It is central to the operational independence of a transmission network that it is able to give assurance of service and maintain itself without any dependence on the higher-layer switching network information. For this, it needs to run an operations, administration, and maintenance (OAM) flow with which to verify that the actual traffic in the network is the traffic the connectivity management/control system thinks is configured. And for the OAM flow to know that the right traffic from the specified sources has arrived at the required destination, and only that right traffic has arrived, the source must be implicitly specified by the port through which the traffic passes out of the transmission network.

In fact, the principle of data transparency when coupled with the need for any transmission network to manage its own reliability means that the transmission network must only support point-to-point or simple, statically replicated point-to-multipoint services.[5] This is illustrated in Fig. 4.

QOS TRANSPARENCY

The principle of QoS transparency means that any switching network should be able to manage

[4] *It can be established using basic information theory that transparent data can be copied the number of times needed to achieve point-to-multipoint transport without information loss to the transparent data [3].*

[5] *It can also be established using basic information theory that any merging of the symbols (e.g., packet payloads) of transparent data will result in information loss. As a result, multipoint-to-point (or multipoint-to-multipoint) must have extra information added to compensate for this information loss through merging. This extra information is exactly that needed to demultiplex the merged information back to point-to-point or point-to-multipoint transport. If a transmission network is to independently monitor its transfer of the transparent data, this extra information must be added by the transmission network itself [3].*

If the transmission network is to be able to reliably offer this QoS with full operational independence from any switching network, the transmission network needs to know that contention within the transmission network is entirely within its own control.

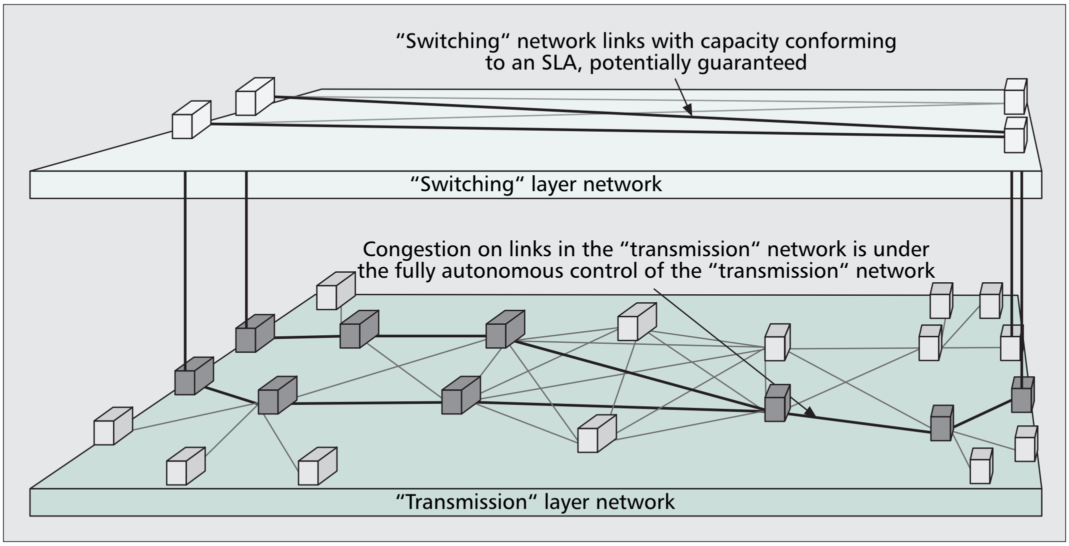


Figure 5. *QoS transparency.*

its own allocation of resource capacity. This means that as long as the switching network meets the basic traffic parameters set by the transmission network ingress port — which the switching network can control — the transmission network can, if appropriate, guarantee to carry the data to the destination(s).

Again, as illustrated in Fig. 5, if the transmission network is to be able to reliably offer this QoS with full operational independence from any switching network, the transmission network needs to know that contention within the transmission network is entirely within its own control. This too results in the practical restriction that the only services the transmission network can offer are point-to-point or simple, statically replicated point-to-multipoint.

So by examining some of the consequences of the functional and operational independence between switching networks and transmission networks, we see that the transmission network must only offer point-to-point or simple, statically replicated point-to-multipoint services, and that all management and control of the transmission network must be based on information generated entirely from within the transmission network.

Requirements to Make Ethernet Technology a Suitable Replacement for SONET/SDH Crossconnects

There are a number of technical obstacles to the reuse of Ethernet equipment technology in carriers' networks. Ethernet bridging has been developed as a local area networking (LAN) technology and has a very strong emphasis on auto-configuration: plug boxes together and they work. However, the very protocols that are highly effective in achieving auto-configuration are not normally scalable to the large networks operated by carriers. In addition, the technical mechanisms for managing small-scale local Ethernet networks will also not scale to large carrier networks.

The key developments needed to redeploy Ethernet equipment technology as a replacement transmission technology for carriers are the following:

- Disabling of the bridge learning protocols (i.e., broadcast unknown and source address learning) and any spanning tree protocol in order to remove the barriers to scaling
- Provision of a "northbound" management interface to a network management system and/or an out-of-band distributed control plane
- Inclusion of IEEE 802.1ah "MACinMAC" encapsulation at the network edge to give full transparency between client MAC space and carrier MAC space
- Inclusion of unicast OAM at the network edge, which is a simple extension and profile of IEEE 802.1ag Ethernet bridging OAM
- Addition of "spotlight" monitoring OAM within the carrier Ethernet network
- Inclusion of protection switching mechanisms coupled to the unicast OAM
- Provision of ingress QoS policing on a per transmission service basis at the network edge

The primary objective is to implement these requirements without losing the advantage of reusing the existing enterprise Ethernet technology. The disabling of the bridge learning protocols can normally be achieved even on existing enterprise equipment — disabling of functionality normally does not require bespoke development. All the other functions can be added as "adjunct" functionality on the side of existing developed technology.

Standardization for all of this work is currently underway in IEEE 802 as "Provider Backbone Bridging — Traffic Engineering (PBB-TE)" under project IEEE 802.1Qay [5, 6]. However, most of the technical details are already available in other specifications, so much of this work is cataloguing these technical requirements and describing how the end-to-end network works. In practice, the specification is very well advanced.

It also turns out that Ethernet also offers a

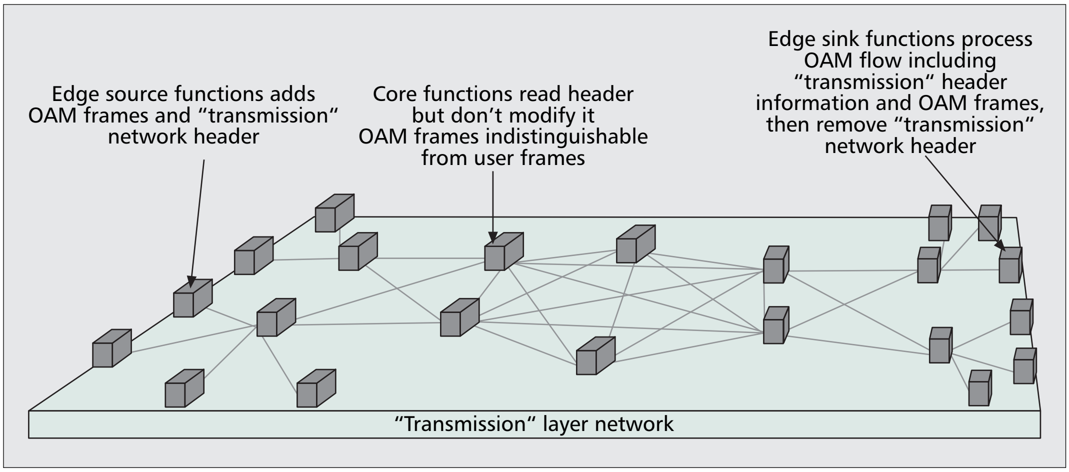


Figure 6. *Forwarding and OAM in PBB-TE.*

While operational experience with these technologies did note practical problems arising from miscofigurations of label swapping after a connection had been up and running for some time, loops were not noted as a practical operational problem.

number of potential technical advantages over other packet technologies for a replacement transmission technology.

First, the Ethernet header is a full and complete header. It is possible to run the whole functionality of the carrier Ethernet network using the header. Any network, in order to operate independently, needs to support a number of information fields that are bound to the service. In packet networks this information is encoded in the header fields.[6] The five critical pieces of information are:

- A destination identifier — encoded as the destination MAC address
- A source identifier — encoded as the source MAC address
- An instance identifier — encoded as a VLAN id
- A route/queuing discriminator — encoded as a VLAN id and/or a priority code
- A service type indicator — encoded as an Ethertype

This means that the Ethernet header can be taken in its standardized form and used for a wide variety of network types. PBB-TE takes this standard header and, by replacing the existing Ethernet connectionless control plane with a connection-oriented control plane, achieves a fully functional independent connection-oriented network.

Second, Ethernet uses an end-to-end header; the values in header fields are not changed as they pass through switches. While changing the value of the forwarding field — label swapping — may give an apparent scalability benefit, the 48-bit MAC address means that Ethernet can achieve sufficient scalability for a transmission network without the need for label swapping.

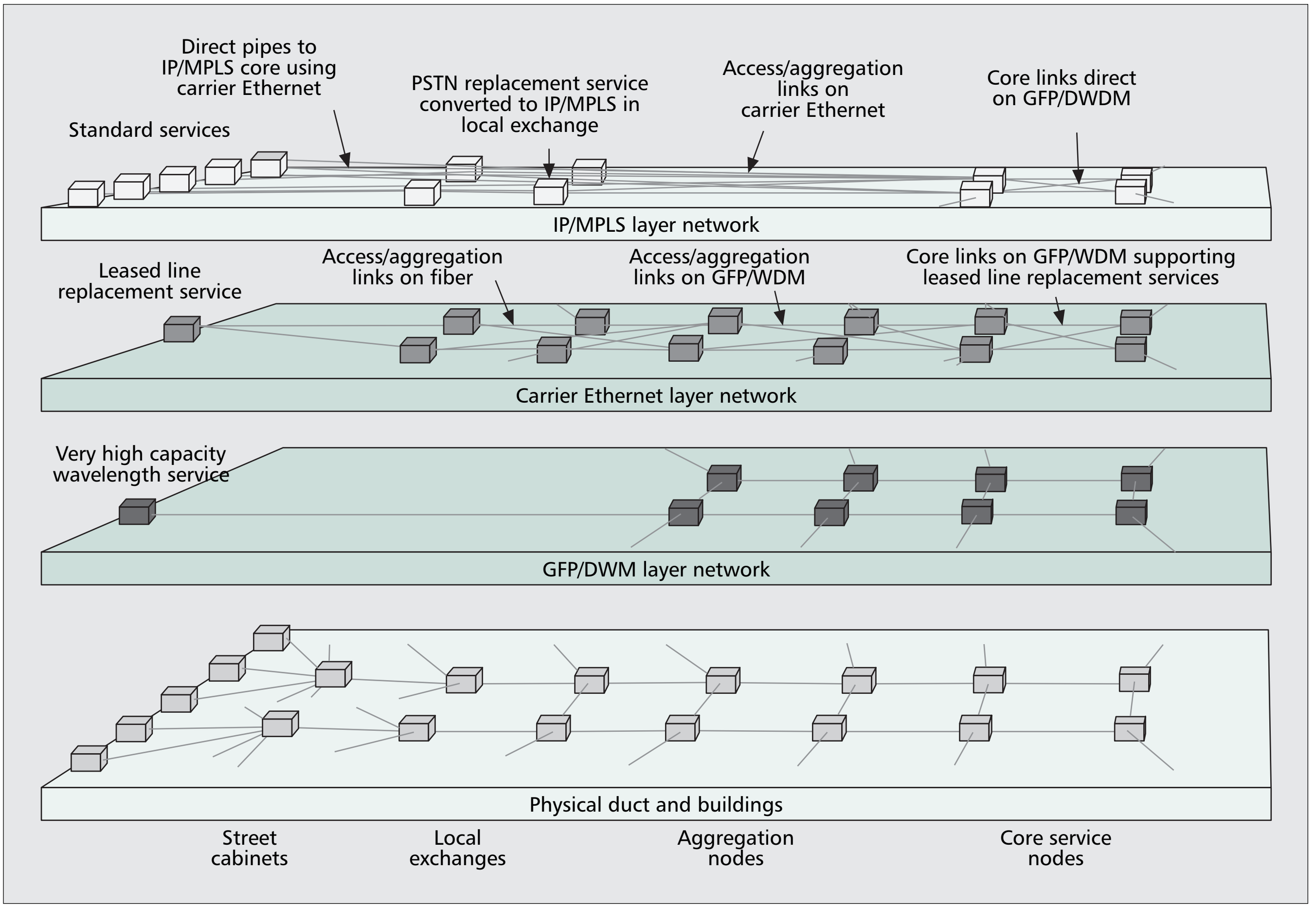
As equipment hardware becomes more reliable, an increasingly important source of unreliability in transmission networks is caused by misconfigurations in the network. If a label is swapped, it requires a configuration, which is an inherent source of unreliability. PBB-TE does not have label swaps and thus is not susceptible to misconfiguration of label swapping. Moreover, the Ethernet header also contains a source MAC address that is also unchanged through the network. This means that every single packet carries an inherent "connectivity verification" check. This makes devising OAM functions much simpler and thus makes the implemented OAM flows more reliable and more cost effective to operate. This is especially important as often the reliability of the OAM flow itself is the weakest link in achieving overall network reliability. PBB-TE forwarding and OAM are illustrated in Fig. 6.

Some have noted that the Ethernet header does not have a time to live (TTL) field and therefore might be susceptible to uncontrolled micro loops. IP and MPLS headers include a TTL field as these technologies include forwarding modes whereby forwarding decisions can be made independently in different switch/routers. When routing information updates at different times in different switch/routers, it is readily possible for packets to forward back to a switch/route through which they have already passed, thereby sending packets continuously around the same loop. The TTL field enables packets in such loops to eventually "die."

With source routed connection orientation, as specified in PBB-TE, this situation does not naturally arise as every end-to-end route is calculated once using one self-consistent set of routing information.

In practice, in this regard PBB-TE is no different from all previous connection-oriented technologies including X.25, ATM, and frame relay. While operational experience with these technologies did note practical problems arising from miscofigurations of label swapping after a connection had been up and running for some time (e.g., traffic suddenly "black-holing" or, worse, being misinserted into some other customer connection), loops were not noted as a practical operational problem.

[6] *These do not always directly map to header fields. For example, the DLCI in frame relay covers the destination identifier, the source identifier, the service instance, and the route discriminator in the one header field. Conversely, more than one header field may be used to assemble one piece of information. For example, an Ethernet VLAN identifier can encode additional information about the destination and source identifiers.*



■ **Figure 7.** *An example of how Ethernet technology can be deployed by a carrier.*

Hybrid Network Structures with IP/MPLS and Carrier Ethernet

While carrier networks are many and varied, the opportunities for carrier Ethernet are likely to have some common features, and there will be some similarities in the way carriers might want to exploit carrier Ethernet technology.

The data plane part of a carrier's network is likely to comprise IP/MPLS, carrier Ethernet, and DWDM technologies all working in concert. The network structure is effectively the choice a carrier makes in placing these technologies in the physical structure of the network and the way in which they are interconnected.

A typical carrier's fixed network using carrier Ethernet technology is illustrated in Fig. 7.

The access and aggregation part tends to be expensive and require long planning lead times to deploy. The use of carrier Ethernet technology in this part of the network is potentially attractive as it provides a cost-effective way of bringing traffic to more central points using a technology that can also outlast current foreseen service requirements.

Service features are supported from more centralized nodes using IP/MPLS technology. The number of central nodes can now be chosen to suit the scope of the services and is unconstrained by the physical structure of the network. In most cases DWDM can be used to transport the Ethernet on the fiber infrastructure.

While there is a simple cost trade-off between the added cost of bringing some traffic up to central nodes and turning it back again, and the savings in capital and operational costs of using carrier Ethernet instead of IP/MPLS in the aggregation nodes, this is not the primary cost advantage. The primary advantage is that the service features can be changed in the future without making any changes to the access and aggregation part of the network.

Between core nodes, IP/MPLS traffic is likely to be quite sufficiently concentrated to warrant transport on direct wavelengths, with carrier Ethernet offering no advantage here.

However, depending on a particular carrier's network, the carrier Ethernet switches themselves may still warrant direct interconnection using wavelengths to support third-party carrier Ethernet services such as leased line replacement services. These services tend to be high-bandwidth and cost-sensitive, and require full data and QoS transparency. In this case they may be economically supported end to end on the carrier Ethernet network without conversion to MPLS. With reasonable consolidation, such an architecture may not result in a significant increase in the number of required core wavelengths, but there may be a simple saving in the amount of IP/MPLS routing required.

CONCLUSION

In this article we have taken a fundamental examination of carrier networks and found there is a clear opportunity for the use of Ethernet technology. In examining carrier networks we find that Ethernet technology, most notably in the form of PBB-TE currently completing standardization as IEEE 802.1Qay, meets the carrier's requirement for a functional replacement for current SONET/SDH crossconnects. PBB-TE has additional advantages as the standard Ethernet header has complete and independent network information, and is unchanged as its crosses the network. These enable PBB-TE Ethernet to be functionally and operationally independent of any traffic it is carrying and to offer reliable, managed, guaranteed service. And these are the features at the heart of mitigating the costs of complexity in carriers' networks.

This strategy of using Ethernet technology works as long as Ethernet technology is kept simple and reliable. There is a risk that some parties will be tempted to add too many service features to Ethernet technology. This would miss the point. Ethernet technology is a highly effective complement to IP/MPLS; it cannot and should not attempt to replace it. Indeed, the great strength of MPLS is that it is closely tied to IP; the great strength of Ethernet technology is that it is not.

REFERENCES

[1] ITU-T Contrib. COM15-C.788, "Complexity, Scalability, Independency, and Transparency", Geneva, Switzherland, Jan. 2008.
[2] ITU-T Rec. G.805, "Generic Functional Architecture of Transport Networks", Geneva, Switzerland, Mar. 2000.
[3] ITU-T Rec. G.800, "Unified Functional Architecture of Transport Networks", ITU-T, Geneva, Switzherland, 2007.
[4] IEEE P802.1ah-2008, "IEEE Standard for Local and Metropolitan Area Networks — Virtual Bridged Local Area Networks — Amendment 6: Provider Backbone Bridges."
[5] IEEE P802.1Qay/D3.0, "IEEE Standard for Local and Metropolitan Area Networks — Virtual Bridged Local Area Networks — Amendment : Provider Backbone Bridge Traffic Engineering", 18 Apr. 2008.
[6] D. Allan *et al.*, "Ethernet as Carrier Transport Infrastructure," *IEEE Commun. Mag.*, vol. 44, no. 2, Feb. 2006, pp. 134–40.

BIOGRAPHIES

ANDY REID (andy.bd.reid@bt.com) is the chief network services strategist in BT's Group CTO and is responsible for defining how network services are supported on BT's network. Prior to this he worked in technical, product portfolio, and regulatory strategy roles. His professional interests include telecommunications technology, traffic management, QoS, functional and enterprise architectures, service pricing, telecommunications microeconomics, and regulatory/competition law economics. He has made major contributions to the standards development of SONET/SDH, and has co-authored two books on the subject.

CHRIS BILTON is director of technology strategy, BT Group CTO. He leads a team that optimizes innovation, technology, and economic analysis supporting the BT LoBs and BT Group. He has over 20 years' experience in creating global networks and services, having held a number of senior roles in BT and BTs JVs. He has a B.Sc. in physics and electronics (Joint Honors) from Brunel University, and is a Chartered Engineer and member of the IET.

IAN HAWKINS is vice president of networks and ICT strategy within BT Group CTO. He leads a team responsible for the creation, communication, and ongoing enhancement of the networks and ICT element of the BT Group technology strategy. The team also provide detailed advice, leadership, and support to BT Lines of Business in areas of technology and service innovation. He has 27 years' experience designing global networks and services, the last 11 in various strategic and technology direction setting roles.

PETER WILLIS (peter.j.willis@bt.com) provides BT's network technology strategy in the BT Group CTO. He led the development of BT's carrier Ethernet requirements and participated in BT's work on IEEE 802.1Qay as part of BT's 21st Century Network Architecture. He has 20 years' experience in distributed computing and data networks, including designing, installing, and operating BT's first public Internet service from 1994 to 1998. He has an M.Eng. in electronic systems engineering from the University of York.