Ethernet nodes in terabit backbones

Heikki Almay
almay@kolumbus.fi

Abstract

Today Ethernet switching is primary used in local area networks and in regional broadband aggregation networks. This paper discusses the potential use of Ethernet and the more recent Carrier Ethernet nodes in terabit backbones now and in the near future. It outlines an evolutionary approach that potentially paves the way to a layer two centric networking paradigm in the backbone as well. Technical and economic aspects of expanding the current high speed router based networks are described.

1 Introduction

Ethernet has long been the dominant technology for local area networking. In the new millennium the technology has been adopted to support Ethernet services over the wide area network. Originally the target of this Metro Ethernet work was to develop service provider's tools for providing cost efficient connectivity for business customers.

Today Ethernet based access networks are widely deployed in various types of broadband aggregation networks and the Ethernet community is boldly looking forward for new application areas including high speed backbone networks.

The transformation of *Switched Ethernet* to *Carrier Ethernet* has not been simple. Ethernet being a flat layer two network does not scale well. It is considered to be difficult to troubleshoot, occasionally generates broadcast storms and is very vulnerable for various types of security threats [2]. As a consequence in Carrier Ethernet key concepts like MAC learning, broadcasting and spanning tree have been dropped. On the other hand also traditional Ethernet switches and switch/routers have been further developed to better meet the needs of service provider networks.

At the moment the key drivers for Ethernet in carrier networks are perceived low cost and the need for capacity increase.

This paper discusses the potential use of Ethernet Switches or Carrier Ethernet nodes in terabit backbones. Instead of proposing a network architecture consisting of switches only a more evolutionary approach is discussed.

This paper seeks to answer the following questions:

- Are Ethernet switches and cross connects likely to replace core routers in the backbone networks the same way ATM replaced Frame Relay and IP replaced ATM?
- What features and capabilities are required from Ethernet switches or cross-connects when they are introduced to the terabit backbone?
- Under what circumstances is it justified to use Ethernet switches or cross-connects as building blocks of a terabit backbone?

In chapter 2 the evolution of backbone technologies is discussed. Link capacities and interface speeds of the different backbone generations are outlined and a brief analysis on the possibilities for Ethernet to achieve a technology shift is performed.

Chapter 3 compares the pricing and cost of Ethernet to other technologies. In chapter 4 the current backbone architecture is outlined and the reader is introduced to an example network that will be discussed throughout the paper. To create some thoughts additionally an example of a perfectly working Ethernet based service connecting high end routers is given.

Chapter 5 takes a dive into the lower protocol layers and suggests the replacement of TDM transport with Ethernet. Quality of Service, fail-over-times, traffic aggregation capabilities and link capacities as well as clocking are discussed.

Finally chapter 6 brings Ethernet nodes to the backbone. The example network introduced earlier is rebuilt using Ethernet switches and alternatively with Carrier Ethernet nodes. Chapter 7 gives some guidance on the use of Ethernet nodes in the backbone.

Last but not least chapter 8 provides some conclusions and ideas that have come up when writing this paper. These might be a good basis for further work.

2 Evolution of backbone technologies

2.1 Capacities of backbone technologies

When looking at the history of data networking Frame Relay was developed to provide a simple and fast data forwarding in high quality backbone environments where bit errors were rare and X.25 with per link acknowledgements and retransmissions was considered too complex and resource consuming.

ATM switches did not replace Frame Relay overnight, but in the role of high speed multiservice nodes with STM-1 or STM-4 interfaces they gradually pushed the Frame Relay switches to the edge of the networks. While it would have been possible to construct high speed frame relay switches it was easier to build switching fabrics using fixed length cells.

The same way frame relay was pushed to the edge of the network ATM switches were in turn replaced by routers with STM-16 or STM-64 interfaces in the backbone. When looking back one may ask why ATM was deployed at all as router capacities surpassed those of ATM switches already in the second half of the 90s. Probably the biggest reason was that still in those days the backbone networks were predominantly operated by traditional telephone companies. The telecommunications industry still viewed ATM as the next generation transfer mode that could be adopted to support all traffic – including IP. [13]

Note that each of the technology shifts was associated with a significant increase in backbone capacities. The next generation technology was introduced when the capacity requirements outgrew the old network. In many cases it would have been possible to build a next generation network with the old technology but instead the next generation was selected as it was considered more *future proof*. When considering the growth in data traffic and especially Internet traffic [1] this has been in most cases a well founded strategy. Figure 1 below shows the maximum interface speeds offered for the different network technologies.

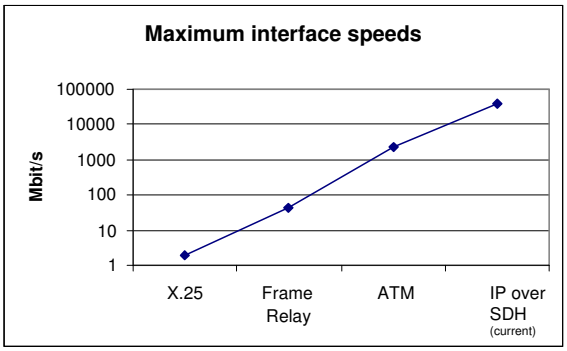

Figure 1: Maximum interface speeds for different technologies

In the figure above the maximum interface speeds for X.25, Frame Relay and ATM are stable as these technologies are considered legacy and there is no commercial interest in higher speed products even though it would be possible to develop some.

Looking at Figure 1 one is tempted to suggest that the next generation backbone technology should support N x 100 Gbit/s interface capacities in order to provide today a similar quantum leap as seen in the earlier technology shifts.

From the earlier technology transitions the assumption can be derived that the next generation backbone technology should provide higher interface speeds than supported in the current backbone router platforms.

At this stage it should be asked if there is a reason to assume that Ethernet node capacities will develop significantly faster than those of IP routers.

A simplistic view on capacity is to look how much traffic each interface card can handle and then count the cards that can reasonably be connected to a system. The first question is discussed in 2.2. The second is more complex as architectural differences (e.g. local switching on interface cards) should be fully understood.

As in both technologies variable size packets have to be switched from one interface to another or alternatively copied to several output interfaces. Consequently there is very little reason to believe why Ethernet switch capacities should in the near future exceed those of backbone routers. When looking at the trend it even seems that the speed at which router capacities are increasing is not slowing down. In just one decade backbone router capacities have jumped from 2Gbit/s to Terabits per second. [12].

2.2 Limitations to interface speeds

For currently deployed IP routers realistic interface speeds are 40 Gbit/s. This corresponds to an STM-256 interface or four 10GE aggregated using 802.3ad or EtherChannel. Both options are commercially available. While some platforms might nominally support a higher number of aggregated 10GE links, the high end platforms today only support 40 Gbit/s connectivity between line cards (e.g. Juniper T640, Cisco 7600, CRS-1). So there are currently no switches or routers that could send more than 40 Gbit/s of user traffic to a link. For the recently announced Juniper T1600 100Gbit/s capacity per slot is claimed.

It is worth noting that the high end platforms listed above are routers. Only the Cisco 7600 should be considered a switch/router. If needed it can be configured to act as an Ethernet Switch or a Carrier Ethernet node.

Current platforms of Ethernet switch vendors have generally lower capacities. This is however not surprising as the products in question are either targeted for corporate markets or to broadband access networks for traffic aggregation.

For Ethernet the 10 Gbit/s interfaces represent the highest bit rate standardized. As mentioned earlier this rate can be multiplied using link aggregation. This solution is however not very useful for wide area connectivity as each 10 Gbit/s interface requires its own wavelength (or fiber pair). When looking at the status of the IEEE 802.3 Higher Speed Study Group new 40 Gbit/s and 100 Gbit/s Ethernet standards are not to be expected before year 2010.

For SDH STM-256 (40 Gbit/s) is the highest bit rate standardized. STM-1024 (160 Gbit/s) is still work in progress. Only very few references to it can be found.

2.3 Ethernet switches and cross connects replacing core routers

Looking at the quantum leaps in capacity that took place when ATM replaced Frame Relay and IP replaced ATM it can be concluded that Ethernet products are unlikely to massively replace routers in the core networks in the next years. There are two key reasons for this.

- For the time being high end router platforms seem to be of higher capacity than high end Ethernet switching platforms. Neither are there any public indications about this changing in the near future. Consequently there is no similar pull from IP to Ethernet as there was from Frame Relay to ATM or from ATM to IP.
- The current high end IP router platforms are utilizing the fastest standardized transmission technologies (STM-256 and 10GE with link aggregation). Any next generation backbone technology would have to rely on yet to be developed standards or proprietary transmission. Both approaches cause at least a temporary obstacle to market entry.

For a network operator the above means that when he is sending out requests for information regarding high speed backbones he will get reasonable answers from both Ethernet vendors and router vendors. Moving away from the current router based architecture to Ethernet only nodes cannot be justified by unavailability of suitable router products, but a second potential reason remains. That is the cost of the products.

3 Cost comparisons

The general perception is that Ethernet is cheaper than competing technologies.

3.1 Interface prices

When looking at *Ethernet transport* it is easy to compare the market prices of Ethernet interface cards to hardware that provides similar transport capacities using PPP or ATM over SDH (or PDH).

The figure below shows the relative price of router interface capacity for a commonly used switch/router. Interface capacity for an STM-1 ATM unit is 300 times more expensive than Gigabit Ethernet LAN capacity and 200 times more expensive than 10 GE LAN capacity. Ethernet WAN interfaces that have better queuing features are far more expensive but still significantly lower priced than SDH based alternatives.

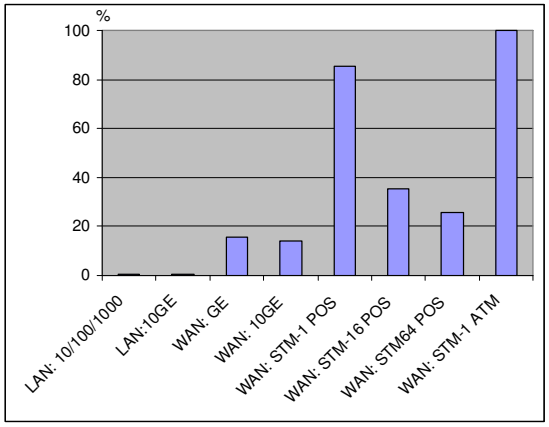

Figure 2: Relative prices of different interface technologies

The figure above is composed using the list prices of one product. So the high capacity cost variations between the different SDH based interfaces can partly be explained by product specifics, but the difference between Ethernet and SDH cannot. Note also that low speed PDH interfaces were not shown in the figure above. The cost of these, especially the commonly used E1 is much higher than for any of the technologies shown.

The conclusion from the above is that capacity in Ethernet transport interfaces is cheaper than alternative interfaces. Especially when LAN interfaces can be used the difference is several orders of magnitude.

Note that the above conclusion is valid for market prices. It does not necessarily reflect the true cost of the different technologies.

3.2 System cost

The claim that Ethernet switching is more cost efficient than routing is more difficult to justify. Obvious major cost differences cannot be identified.

The forwarding tables needed for switching and routing are similar [provided that no huge MAC tables for a flat layer two network are needed].

When compared to a traditional Ethernet switch the router needs additional processing power for running the routing protocols and for calculating the routing tables as well as a big enough memory to store the routing information. The benefit brought along with these additional resources is that router networks scale much better than traditional switched Ethernet that suffer from limited size MAC tables and the growing broadcast and spanning tree domain.

In Carrier Ethernet the automated control plane is switched off and has to be implemented in external server platforms or by skilled network managers doing the work manually. This has a cost. So when considering equipment and service provisioning Carrier Ethernet should be more expensive than traditional switched Ethernet.

A price study comparing routers and switch/routers in a rather similar way as done in section 3.1 can be found at [3]. No major cost differences between switch/router and router platforms are identified but instead the assumption is that the difference is in vendor pricing.

As using public information it is not possible to further analyze the costs of different types of platforms in this paper the assumption is that Ethernet nodes are lower priced than routers and thus preferred [at least by a large segment of network operators]. At the same time it should be kept in mind that in a tougher competitive environment router vendors do have the possibility to lower their prices to match the switch based competition. This assumption is well supported by the fact that the key router vendors report extremely high gross margins on product sales.

4 Backbone architecture

4.1 Backbone structure in carrier networks

In the access network the cost per bit transported is usually the most important issue when network technologies and topologies are selected.

In the backbone also resilience and scalability are important. Often the target is to reach 99.999% service availability. This equals an average of three minutes per year of service downtime. Applications may also require very fast failover times. Consequently multihoming and duplicated equipment are standard design practices.

Still today extensive engineering is required to meet sub-second failover times in case of router failures. Additionally only the latest router platforms provide even rudimentary in service software upgrade capabilities. Consequently most service providers have designed their backbone networks using a two chassis approach.

A general [small] backbone design is shown in the figure below. On the sites hosting the service machinery a pair of provider edge (PE) routers connects to the backbone. In the figure the provider (P) routers are divided into two independent planes. These are not connected to each other. Note that P devices are only present on some of

the sites. Often PE routers of more remote sites connect to the P routers over a wide area link.

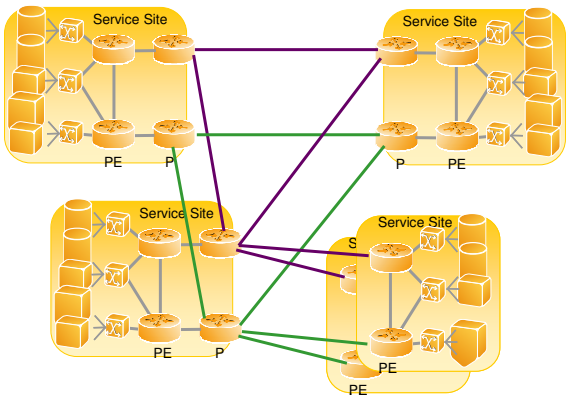

Figure 3: Backbone with two planes

The dual plane architecture is very resilient. However in case an active PE device or the link between the PE and P device fails the problem has to be noticed at the other PE devices so that traffic can be routed to the other backbone plane.

As an alternative to the dual plane architecture a mesh structure can be used. The mesh structure allows capacity optimization and a local resolution of the PE-P link failure but generally the behavior and interdependencies in a mesh are difficult to analyze.

4.2 Use of Ethernet between backbone networks

Internet exchange (IX) points provide a good example of a large volume Ethernet service that is comparable to a backbone. While it is also possible to implement IX services as a routed service or using ATM most implementations use two physically separated Ethernet switches.

All customers connect to the two physically separated switches of the IX provider. This is outlined in the figure below. Each ISP is responsible for his own BGP configurations. All parties can so choose with whom they want to exchange traffic.

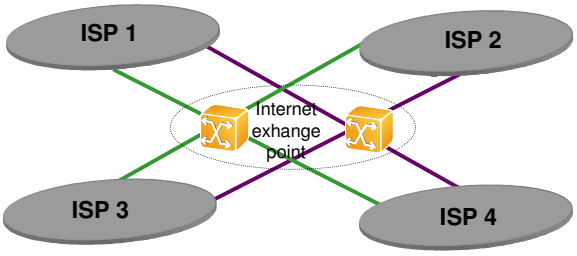

Figure 4: Internet exchange point implemented with Ethernet

In this setup the connectivity between the Internet Services Providers is implemented using standard LAN technology. VLAN traffic separation can be used for differentiating between different protocols for which peering is provided (e.g. IPv4, IPv6, MPLS). For more details on IX implementation see e.g. [4].

Note that the IX service does not make use of any Carrier Ethernet features. Resilience is implemented on the IP layer.

A short answer to the question "*What features and capabilities are required from Ethernet switches or cross-connects when they are introduced to the terabit backbone?"* is none. A pair of big enough Ethernet switches will do. To what extent the layer two IX model can be applied in larger backbones is discussed in the next chapters.

5 Ethernet replacing SDH

5.1 Protocol stack options

The figure below shows the most common protocol stack options for carrying IP traffic in a fiber network.

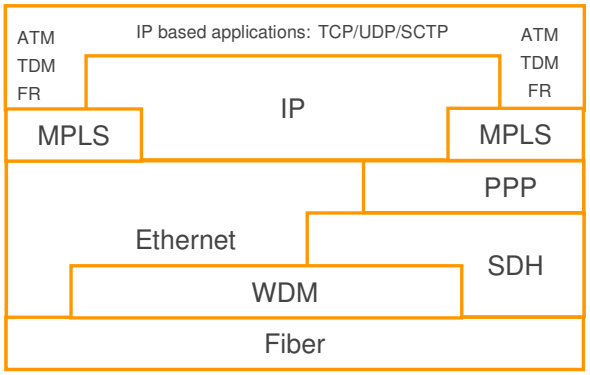

Figure 5: Protocol options for carrying IP traffic in a fiber based network

As discussed earlier Ethernet interfaces are very cost efficient when compared to SDH interfaces in routers. When looking at the different options for constructing the protocol variants Ethernet over WDM has become a significant alternative for the traditional packet over SDH.

When Ethernet transport is used as an alternative to SDH it has to be analyzed what SDH functions can be implemented by Ethernet and to what extent those that cannot can be absorbed by other protocol layers.

5.2 Quality of Service

From the IP layer perspective SDH provides high quality transport for all traffic. In a TDM system delay and latency are minimal. High availability and resilience mechanisms are available and 50 ms failover times can be guaranteed. Additionally SDH provides aggregation of traffic to high speed links and cross connects where needed.

Ethernet nodes support at least IEEE 802.1p frame priority for quality of service differentiation. Many products also support queuing based on IP diffserv codepoints. So with Ethernet more urgent traffic can be priorised, but limiting packet loss and delay variation requires careful network planning [or heavy overprovisioning]. This has to be kept in mind when the traffic aggregation capabilities are discussed in chapter 5.4.

SDH systems also perform bit-error rate monitoring and provide alarms for both degraded signal and failed signal. Such information can be used for both routing and maintenance purposes. Comparable features are currently missing from Ethernet products.

5.3 Service availability and fail-over times

With Rapid Spanning Tree (RSTP) and Multiple Spanning Tree (MSTP) sub-second failover times can be reached. In a Carrier Ethernet environment the vendor specific Ethernet Ring Protection schemes can be used. With these mechanisms failover times are comparable to those of SDH. Note that it depends on the network structure and the configuration which of the resilience and loop protection mechanisms discussed above are applicable.

When IP is run on top of an SDH or PDH transmission system, problems of the transmission network are signaled to both ends of the connection within milliseconds. Today routers can utilize this information for re-routing traffic to alternative links. In Ethernet networks information about a failure remains local. If two routers A and B are connected to each other using an Ethernet switch each of the routers will notice if the local Ethernet link will go down. A problem in the link between router A and the switch will not be noticed by router B which will continue to send traffic to the link until the used routing protocol will eventually notice that router A is no more reachable using the link in question. In order to speed up the failure detection routers can use bidirectional forwarding detection (BFD) in order to speed up the discovery of connectivity problems that are not noticed by the line cards. Using BFD failover times of 150 – 200 ms can be achieved.

Note that several standardization bodies are working on improving the Ethernet management capabilities. For the backbone probably the most interesting is 802.1ag (Connectivity Fault Management). It will provide tools for practical troubleshooting of Ethernet connections i.e. Virtual Circuit Connectivity Verification, Ping, and Traceroute.

5.4 Traffic aggregation

Similar to SDH Ethernet switches can aggregate traffic to high speed links and provide cross-connect (or switching) functionality. A clear benefit of Ethernet [and any other packet data service] is that capacity utilization for data services is much better than in TDM systems as no bandwidth has to be reserved for the individual connections. In SDH networks a fixed bit rate has to be allocated for each connection. As capacity is typically reserved to carry the peak rate traffic a significant portion of the capacity remains unused most of the time. If the typical ring protection schemes are used the situation is even worse. In addition to the underutilized active link a similar unused link is provisioned. Often service providers avoid this problem by resorting to layer three resilience and use the two alternative SDH links in parallel.

Note that while Ethernet is efficient for data traffic with long packets, it can carry voice only with a significant overhead.

5.5 Link capacity

When looking at the nominal bit rates until recently SDH has been the *high capacity* technology. STM-64 (10 Gbit/s) backbone networks have been operational since the late 1990's. The 10 Gigabit Ethernet standards only emerged half a decade after that. Traditionally the capacity per wavelength or (pair of fibers) has been much better in SDH than in Ethernet. Today there is parity and while STM-64 is still more widely deployed 10 GE has become an accepted alternative.

The STM-256 (40 Gbit/s) specifications have been available since year 2000 and a small scale commercial STM-256 market exists since 2005. 100 GE is still in standardization and expected to be ready in 2010.

5.6 Clocking

Another issue important to several applications is clocking. SDH is traditionally used for clock distribution in large telecommunication networks. At the moment clocking cannot be delivered by Ethernet itself although first synchronous Ethernet products are entering the market. Generally timing over packet networks is provided using other means e.g. using external clock sources or adaptive clocking.

5.7 Replacing SDH with Ethernet

Key issues to take into consideration when SDH is replaced by Ethernet in the core network are:

- Current Ethernet resilience schemes are not as fast as those implemented in the SDH network. In stead of tuning the available L2 methods it is often easier to resort to L3 capabilities (routing protocols or BFD)
- Quality of service requires special attention especially if heavy statistical multiplexing is used for increasing network efficiency. Also bit error monitoring should be made available.
- Some applications may require clocking to be distributed by the transport network. For such cases several schemes for timing over packet are available. Alternatively external clock sources (e.g. GPS receivers) can be used.

While the above differences make a transition from SDH based transport to Ethernet transport in the backbone more complex the cost benefits and increased efficiency for data traffic suggest that within a few years Ethernet will become the dominant technology for new wide area transport provided that the issues listed above are resolved an that the added complexity will not fragment the Ethernet market and remove the cost benefit that has its roots in high volume products.

Note that even though Ethernet is on its way to become the dominant technology for new projects the existing transport networks represent a huge asset for the network operators. These will exist and be expanded for many years to come. Especially in lower capacity networks where SDH is operated directly over fiber the installed transmission systems may cause a severe obstacle for deploying an Ethernet only infrastructure.

6 Terabit Backbone with Ethernet nodes

6.1 Example network

Let's go back to the network shown in Figure 2 and rebuild the example network using Ethernet nodes. The new network design is shown in the figure below. Note that now the sites are named A, B, C, D and E.

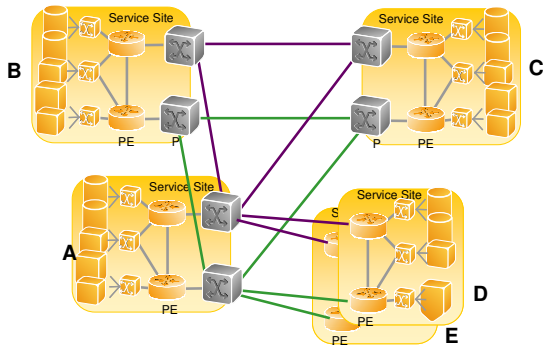

Figure 6: Backbone with Ethernet Switches

In order to shed light to the question *what features and capabilities are required from Ethernet switches or cross-connects when they are introduced to the terabit backbone* we can start by analyzing how the network would work if the same type of layer two service was provided by the backbone switches as in the IX example in section 4.2. After that we can analyze how the network changes if Carrier Ethernet products are used and the traditional Ethernet control plane is disabled.

6.2 Building the network with switched Ethernet

The use of Ethernet nodes affects the design of all the three lowest OSI layers. With Ethernet nodes the options for the physical wide area transport are limited to Ethernet over WDM (or directly over dark fiber) and NG-SDH. This is however in most cases not an issue in the backbone area.

In the Ethernet design the first critical issue is loop protection. Sites A, B and C form a ring – or more precisely the two backbone planes form each a ring. For each VLAN the ring has to be cut at some point. The straightforward solution is to build the network with enough capacity and just let spanning tree cut the rings somewhere.

If surplus transport capacity is not available or delay and jitter should be optimized it is desirable to configure the network so that traffic between C and the sites A, D and

E is not carried via B. On the other hand traffic from B to A, D and E should under normal circumstances not be carried via B. It is also desirable that traffic between B and C flows directly. The three different topologies can be achieved using three VLAN with different MSTP instances. For the example network this is probably a good and reliable solution. It should however be kept in mine that the spanning tree protocols do not scale well. The addition of a node to the network or any topology change will lead to a consecutive spanning of all the links in the STP domain.

The resilience could also be implemented on the IP layer. This would however require a loop-free Ethernet topology. So on both backbone planes the ring should be cut. This [as discussed above] does not allow optimal use of transport capacity.

In the new setup with no routed backbone the PE routers all are neighbors to each other. In a rather small network as used in the example this is not a concern. With five sites and a dual plane architecture each PE router exchanges routing information with four other PE devices. Normally when MPLS L3 VPN services are enabled the PE routers exchange anyway VPN route information with each other using BGP. For scalability route reflectors can be added [7]. Also the processor load on routers caused by BGP has been studied [8]. Even during abnormal events [like the SQL Slammer worm] CPU load is not likely to increase significantly. Note that in case of a traditional IP/MPLS backbone the PE devices would participate in the [typically IS-IS] routing inside the MPLS core.

From the perspective of the PE routers the Ethernet switch based backbone means that no information about link or node failures in the backbone will be reported back or forward. BFD or an alternative protocol has to be configured between the routers for fast detection of failures. Some years back this approach would have caused a heavy load on the central route processor of a router but more recent router models can handle BFD on line cards. So moving to Ethernet should have only a minor impact on the fail-over times.

Transmission capacities required on the backbone links grow when moving from PPP to Ethernet. PPP overhead per packet is 8 octets while Ethernet II frames cause at least a 24 octet overhead assuming full duplex operation. If the bulk of the traffic consists of long packets the difference is neglectable, but e.g. for mobile voice where the size of the IP packet is around 80 octets, the added L2 capacity need has to be considered in network planning.

Where the service provider may have an issue is when service quality in the switched backbone is degraded but still available. In such cases (increased bit error rate, locally misconfigured max. MTU size etc.) BFD and routing protocols still work but some user applications might not.

Another potential source of problems is related to broadcast. Broadcast traffic cannot be fully disabled as the network elements find each other using ARP requests that are flooded in the VLAN networks to which the network elements in question belong. Major switch vendors provide features and guidance for [broadcast] storm control. While the example network is small the broadcast challenges limits the scalability of a standard LAN as a terabit backbone.

From the above discussion it can be concluded that the example network as shown in Figure 6 can be implemented using current high end LAN switches provided that a fiber network or proper WDM systems are available. For ensuring proper behavior in case of Ethernet network problems the PE routers need a scalable BFD implementation. They also have to be resilient and behave in a deterministic way when facing high amounts of broadcast traffic. Additional functionality for troubleshooting degraded connections would be desirable.

Resolving the scalability and operational concerns listed above has been one of the key targets of Carrier Ethernet.

6.3 Selecting the right flavor of Carrier Ethernet

Carrier Ethernet comes in many flavors.
Plane IEEE 802.1q allows MAC address learning to be disabled for point-to-point VLANs in a bridge [10]. In our example network it would be very straightforward to configure the connections from a PE device to all the others as VLANs (virtual interfaces from the router perspective). In the Ethernet backbone with N sites this would result in Nx(N-1)/2 point-to-point VLANs across the network. In the example with five sites ten VLANs per plane would have to be set up. The two challenges with this approach are that it is not widely supported and the problem of scaling. If N grows to 90 the number of needed VLANs is 4005 and the provider is running out of VLAN identifiers. If more than one VLAN has to be carried between two sites the ceiling is hit with a lower number of sites.

The second alternative is to use Provider Bridges (Q-in-Q, IEEE 802.1ad). It scales the same way as 802.1q but in this case carrying several VLANs from one site to another site would not affect the scaling of the network.

For overcoming the scaling problems Provider Backbone Bridge. (PBB) also known as MAC-in-MAC and IEEE 802.1ah was developed.

All the above Ethernet variants still basically rely on MAC learning, broadcast and spanning tree. All these are removed in Provider Backbone Bridge Traffic Engineering (PBB-TE) also known as Provider Backbone Transport, IEEE 802.1Qay. The specification of PBB-TE is currently ongoing.

A further alternative would be the use of VLAN Cross Connect as proposed in [10]. It is a method to establish dedicated point-to-point connections through the Ethernet network and to overcome the scaling problem by restricting the significance of a VLAN identifier to a link. Using this method, frames are forwarded according to the ingress VLAN identifiers that appear in the frame. This approach is at the moment not progressing in standardization and the feature is not widely supported.

From the different options discussed above PBB-TE seems to be commercially and from standards perspective the most promising approach. The disabling of the traditional Ethernet control plane removes some of the concerns discussed in section 6.2 for switched Ethernet.

Consequently below the example network is discussed with primarily PBB-TE in mind. Note that the discussion is based mostly based on expectations and interpretation on how the network should work when built according to specifications. Only limited and predominantly theoretical information about the actual performance and operation of Carrier Ethernet networks is available. [14]

Current Carrier Ethernet implementations are mostly used for aggregating DSL traffic

6.4 Building the network with Carrier Ethernet

For Carrier Ethernet the same general constraints apply as to switched Ethernet. PE routers all are neighbors to each other. Wide area transport is limited to WDM, dark fiber and NG-SDH. QoS schemes are the same and link quality monitoring capabilities are equally missing although ITU-T is working on Ethernet Operation and Maintenance including Performance Monitoring.

Moving from traditional Ethernet to Carrier Ethernet does not change the loop prevention issue in the example network. In stead of spanning tree protocols vendor specific Ethernet Ring Protection schemes e.g. [9] could be used. The Ethernet Ring Protection promises similar failover times as SDH. In practice 50-100ms are realistic. Note however that in a dual plane network [as in the example in Figure 6] where static paths between the tunnel endpoints are configured additional protection may not be needed. Resilience can be conveniently implemented on layer 3.

The additional transmission overhead caused by Carrier Ethernet (PBB-TE) per packet is double when compared to traditional Switched Ethernet. Again for large packets and a fiber network this is not an issue, but for network planners facing narrow links carrying voice traffic this may be a reason to select an alternative technology.

As the PE routers generally support VLANs it is possible to configure paths to multiple destinations from a single PE router port. In a backbone application this is especially valuable as the traffic pattern is a mesh between the sites. In the example with five sites ten paths per plane have to be set up.

An additional aspect that should be considered when moving towards Carrier Ethernet is the provisioning and management of the connections. Connections have to be set up manually or using a [yet to be developed] automated control plane, e.g. GMPLS. In the example network each connection only traverses two Ethernet nodes. In a large network this figure would be higher and so also the configuration effort per connection.

The centralized manual or automated provisioning scheme may cause some resilience concerns. Switched Ethernet as well as IP/MPLS automatically adopt to major topology changes such as the loss of a major site with several nodes. For Carrier Ethernet the first question is what part of the network is still accessible for management. Disaster recovery of the management system itself should also be addressed. Also more local problems like overload conditions can jeopardize the manageability of the network.

From the above it can be concluded that the network shown in Figure 6 can be implemented using Carrier Ethernet. In a larger terabit backbone the Carrier Ethernet development addresses key issues restricting the scalability of switched Ethernet. Spanning tree and broadcast are disabled. It seems however that in the process a few new challenges are created. The network configuration effort should not be underestimated. Also the challenges related to the resilience of the management and the management connections in case of network problems may cause some headache. Also little evidence is available yet to prove that Carrier Ethernet products do provide working solutions for terabit backbones. The focus of the Carrier Ethernet development and deployments is still in the aggregation networks. Another aspect that should not be forgotten is that some of the same constraints that limit the application of switched Ethernet in wide area networking also apply to Carrier Ethernet. The transmission network

should provide Ethernet or direct optical interfaces and the number of adjacent PE routers should not grow too big.

7 Using Ethernet nodes as building blocks of terabit backbones

Based on the discussion in the previous chapters it is time to ask *under what circumstances it is justified to use Ethernet switches or cross-connects as building blocks of a terabit backbone*.

As discussed in section 6 Ethernet switches and cross-connects can be used in stead of backbone routers when WDM, dark fiber or NG-SDH is available for Ethernet wide area transport. The scalability of switched Ethernet is limited. This may improve with Carrier Ethernet. Also an eye should be kept on the number of adjacent routers and the additional requirement (e.g. BFD) that an Ethernet backbone sets to the PE routers.

The above technical conditions are most likely met when existing backbones are of an established wireline service provider is extended. High capacity needs, own fiber and WDM equipment and technical skills to work out new configurations and operational procedures are available. In stead of deploying a new pair of routers the service provider would deploy a pair of switches.

Also small greenfield backbone projects could be implemented using an Ethernet only backbone. In such a cases it however likely that the operator is relying on transport capacity from other players. Often also external competences are used. So introducing a new technology might be considered as too risky.

In addition to the technical constraints the economic aspects have to be considered in the technology choice. As discussed in section 3 the price per bit for Ethernet transport interfaces is significantly lower than for SDH. The cost differences between switching and routing [or doing both in the same platform] remains unclear and eventually vendor pricing decisions will show if there is a market for Carrier Ethernet in terabit backbones.

When looking from commercial perspective at the potential types of network projects where Ethernet nodes could appear as alternatives to backbone routers the most likely candidates are upgrades or new networks of limited size – in practice cases where both switched and Carrier Ethernet could be applied.

8 Conclusions and further study items

The target of this paper was to elaborate the following questions:

- Are Ethernet switches and cross connects likely to replace core routers in the backbone networks the same way ATM replaced Frame Relay and IP replaced ATM?
- What features and capabilities are required from Ethernet switches or cross-connects when they are introduced to the terabit backbone?
- Under what circumstances is it justified to use Ethernet switches or cross-connects as building blocks of a terabit backbone?

A wholesale replacement of IP routers by Ethernet switches seems unlikely especially as the driver for development seems to be high speed Ethernet interfaces. The most powerful platforms to house these new interfaces are routers or switch/routers. In earlier technology shifts the new technology has provided significantly higher capacity interfaces than the old one.

Then again when comparing Ethernet and SDH it seems very likely that Ethernet is going to replace SDH as the standard transport and cross-connect technology. The transitions in transport networks are very slow and it may be that the existing transmission systems will cause a severe obstacle for Ethernet in the backbones. This concern is however more relevant to the lower capacity networks and not the real terabit backbones.

With some constraints the current high end Ethernet switches with can be used in high speed backbones to replace and to complement backbone routers. Scalability issues related to spanning tree, flat addressing and broadcasts have been addressed in the Carrier Ethernet development, but in the process some new challenges especially in provisioning and management have been created.

Ethernet feature deficits like slower fail-over, missing clock distribution, and the lack of link quality monitoring are mostly due to the missing TDM transport layer.

While the missing features clearly cause new requirements for the routers and this paper fails to set absolute scalability limits for a network with PE routers and an Ethernet backbone it is safe to say that networks with some tens of PE routers can be rather easily connected using Ethernet only.

When looking at *Figure 1: Maximum interface speeds for different technologies* and *Figure 2: Relative prices of different interface technologies* it is tempting to refer

to the work of Clayton Christensen [11] and to conclude that the shift from IP over SDH to Ethernet is like any other technology transition with seemingly inferior next generation [Ethernet] products entering the low end [access] market while the complacent incumbent [router] vendors still enjoy good margins and growth in their core [backbone] business. On the other hand the current lack of next generation interface standards, the position of the incumbent router vendors as the likely source for the highest capacity future platforms and the absence of any proof that switching is cheaper to implement than routing give reason to think twice. Besides it seems to be the SDH market that is to be replaced by Ethernet and WDM. So this work should be continued with a study on *technology transitions in multilayer networks.*

References

[1] University of Minnesota, Minnesota Internet Traffic Studies (MINTS), http://www.dtc.umn.edu/mints/home.html

[2] Paggen, Christopher : Understanding and Preventing Layer 2 Attacks in the Campus Networks, Cisco Networkers 2007, http://www28.cplan.com/sb156/login.jsp

[3] Siegel, David: blog 2.3.2007: The session wasn't named 10Gig and the Next-Gen network, http://blogs.globalcrossing.com/comment/reply/305

[4] http://www.ficix.fi/

[5] http://www.yersinia.net/index.htm

[6] IEEE 802.3 Higher Speed Study Group, September 2007 Interim Meeting http://grouper.ieee.org/groups/802/3/hssg/public/sept07/index.html

[7] RFC 2547, BGP/MPLS VPNs, March 1999 http://www.ietf.org/rfc/rfc2547.txt

[8] Agarwal et al: Impact of BGP Dynamics on Router CPU Utilization, http://research.microsoft.com/~sagarwal/pam04.pdf

[9] RFC 3619, Extreme Networks' Ethernet Automatic Protection Switching (EAPS)Version 1 ,http://www.ietf.org/rfc/rfc3619.txt

[10] Sprecher, N.,Klein P., Lingyuan F., Berechya D.: Internet-Draft, GMPLS Control of Ethernet VLAN Cross Connect Switches February 8, 2006, http://www.watersprings.org/pub/id/draft-sprecher-gels-ethernet-vlan-xc-00.txt

[11] Christensen, Clayton: The innovator's challenge: understanding the influence of market environment on processes of technology development in the rigid disk drive industry, Harvard University, Graduate School of Business Administration, 1992.

[12] McKeown, Nick: Growth in Router Capacity, IPAM, Lake Arrowhead, October 2003, http://tiny-tera.stanford.edu/~nickm/talks/IPAM_Oct_2003.ppt#492,6,Backbone router capacity

[13] Steinberg Steve: Netheads vs Bellheads, Wired, Issue 4.10, Oct 1996 http://www.wired.com/wired/archive/4.10/atm.html?topic=&topic_set=

[14] van Malenstein, G.A., Steenbeek, C.: PBT Networking, Host-to-host connections through SURFnet6, July 2, 2007, http://staff.science.uva.nl/~delaat/sne-2006-2007/p34/report.pdf