On Reliable and Cost-Efficient Design of Carrier-Grade Ethernet in a Multi-Line Rate Network under Transmission-Range Constraints

Marwan Batayneh†, Dominic A. Schupke‡, Marco Hoffmann‡, Andreas Kirstaedter‡, and Biswanath Mukherjee†
† *University of California, Davis*
mmbatayneh@ucdavis.edu, mukherje@cs.ucdavis.edu
‡ *Siemens Networks GmbH & Co KG, Munich, Germany*
{dominic.schupke, marco.hoffmann, andreas.kirstaedter}@siemens.com

Abstract: We study cost-efficient design of reliable Next-Generation Carrier-Grade Ethernet under different link rates and signal-transmission-range constraints. Experimental results from our proposed algorithms show significant improvement on the network cost compared to other designs.

©2007 Optical Society of America

OCIS codes: (060.4250) Networks

1. Introduction

Being the dominating technology for Local Area Networks (LAN), around 90% of Internet traffic is generated by end systems with Ethernet interfaces. Hence, efforts for extending the usage of Ethernet beyond LANs to Metropolitan Area Networks (MAN) or even to Wide Area Networks (WAN) are in progress. The future mode of operation is to carry native Ethernet frames directly over Wavelength-Division Multiplexing based optical backbone networks (Ethernet-over-WDM). Thus, several layers of other technologies can be eliminated, and significant savings in Capital Expenditures (CapEx) and Operational Expenditures (OpEx) can be achieved.

In a backbone network, Ethernet can be set up as a connection-oriented service, i.e., tunnels carrying Ethernet frames. Three possible forwarding technologies for carrying Ethernet frames over such tunnels are being explored: Virtual-LAN Crossconnect (VLAN-XC), Provider Backbone Transport (PBT), and Transport Multi-protocol Label Switching (T-MPLS) [1]. This study [1] found that maximum CapEx and OpEx savings can be achieved by running Ethernet over WDM tunnels (i.e., lightpaths [2]) at high rates (up to 100 Gbit/s).

Our study focuses on Ethernet-over-WDM inside the carrier's backbone (carrier-grade) network, so Ethernet now requires carrier-level reliability. Also, since the lightpath rate can be high (100 Gbit/s) and it may travel long distances (for backbone links), 3R (Re-amplification, Retiming, and Reshaping) signal regeneration may be needed for a lightpath. Thus, we use the term *Etherpath (EP)* to refer to a lightpath carrying such Ethernet connections (or tunnels).

Note that, even though this study is on Carrier-Grade Ethernet, the general concepts and solutions can be applied to connections carrying traffic other than Ethernet frames. In addition, the terms 3R regeneration and regeneration are used interchangeably.

2. Node Architecture and Problem Formulation

A) Node Architecture**:** Figure 1 shows the Ethernet-over-WDM switch architecture. It has two components: (1) Optical Crossconnect (OXC) which performs switching at the Etherpath level and (2) Ethernet Switch (ES). The Ethernet Switch initiates and terminates the Etherpaths using the Ethernet Interfaces (EI), namely Ethernet Transmitter (ET) and Ethernet Receiver (ER). The Ethernet switch can perform electronic functions of grooming Ethernet connections onto the Etherpaths, and regeneration. The OXC must have enough I/O ports (linecards) to support fiber connections with neighboring nodes as well as for local add/drop with the Ethernet Switch.

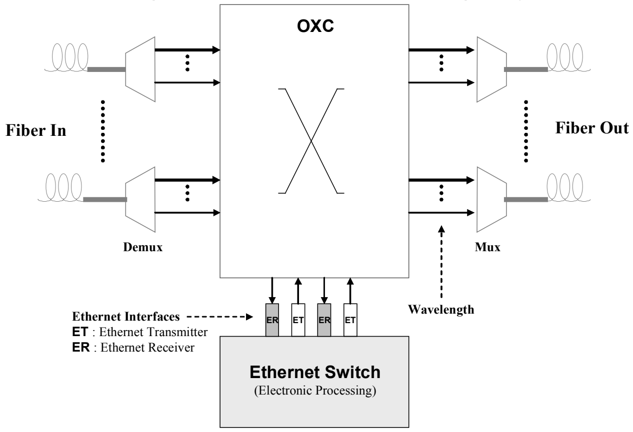

Figure 1: Node architecture.

B) Problem Formulation:

Given: ***1)*** Network's physical topology represented by graph *G(V,E)* with set of edges *E* representing links and set of vertices *V* representing the nodes. ***2)*** Traffic demand matrix. ***3)*** All nodes have regeneration and traffic grooming capabilities. ***4)*** Mixed line rates, i.e., wavelengths on different links may operate at different rates (10/100 Gbit/s). (All wavelengths on a link operate at the same rate.)

Transmission Constraints: ***1)*** Number of wavelengths on each link and wavelength capacity. ***2)*** Node capacity, which is determined by the number of Ethernet interfaces. ***3)*** Transmission range constraints. ***4)*** Wavelength continuity constraint is assumed on the Etherpaths.

Need to: ***1)*** Route all the Ethernet connections. ***2)*** All Ethernet connections must be protected. Connection-Level Shared-Protection scheme is used to protect the Ethernet connections against single link failures [3].

Objectives**:** ***1)*** Satisfy all the Ethernet connections. ***2)*** Reduce the solution's CapEx. In this study, network's CapEx is dominated by the amount of electronic processing required, which determines the number of Ethernet interfaces used.

3. Proposed Algorithms

An efficient algorithm for this problem must address the following observations:

(1) Shortest path, where edge weight is the link's actual physical length, does not imply that the path is a minimum-regeneration path (which is a path that requires the minimum amount of 3R regeneration).

(2) Establishing an Etherpath over a set of links running at rate *R* implies that the transmitter and receiver at the ends of the Etherpath must operate at a *minimum* rate of *R*.

(3) If an Etherpath originating from node X is destined to node Z (EP X-Z) and requires regeneration at a third node Y, then since the EP must be terminated in the ES at node Y to perform regeneration, EP X-Z is segmented into two Etherpaths, namely, EP X-Y and EP Y-

Z. Note that, if EP Y-Z is not fully utilized, then this will create a grooming opportunity (for other traffic) onto EP Y-Z at node Y.
(4) If an Etherpath originating at node X is destined to node Z (EP X-Z) using links X-Y and Y-Z, where link X-Y is running at rate *R1* and link Y-Z is running at a different rate *R2*, then EP X-Z must be segmented into two Etherpaths, namely, EP X-Y and EP Y-Z. EP X-Y runs at *R1*, and EP Y-Z runs at *R2*. Again, if EP Y-Z is not fully utilized, there is a grooming opportunity onto EP Y-Z at node Y.
(5) Once an Etherpath is terminated in the ES, electronic functions are accessible. For example, grooming can be performed, and wavelength conversion is also possible. In addition, regeneration is *always* performed.

To address these issues, we construct a graph called Mixed Topology (MT). The term Mixed is used since the edges of the graph might be physical links, virtual links (already established Etherpaths), or both. The MT is constructed on a per-Ethernet connection basis, i.e., for each connection to be routed, MT is constructed. The MT contains all possible virtual and physical links over which a specific connection can be routed (*any path in the MT is a viable route)*. The MT reflects a partial picture of the network's current state and resource utilization. Establishing the MT for finding a protection path must address resource-sharing constraints. Hence, risk information must be captured. Figure 2 shows the rules for constructing the MT for both working and protection paths. Below, we discuss the design algorithms.

Constructing MT for the Working Path

The MT includes all the physical and virtual links except:

1) All links running at rates below the connection's rate.
2) All EPs reserved for protection.
3) All saturated EPs or EPs with free capacity below the connection's rate.
4) All saturated links (saturated = all wavelengths are used).
5) A physical link or a physical segment (set of adjacent physical links) with both end nodes connected by a working EP is replaced by that EP.

Constructing MT for the Protection Path

The MT includes all the physical and virtual links except:

1) All links running at rates below the connection's rate.
2) All physical links on the connection's primary path.
3) Any protection EP for connections sharing the same risk with the current connection if the aggregate capacity of the connections exceeds the EP capacity.
4) All working EPs.
5) All protection EPs that have reached the maximum allowed sharing, if any.
6) All links saturated with working traffic.
7) A physical link or a physical segment (set of adjacent physical links) with both end nodes connected by a working EP is replaced by the EP.

Figure 2: Constructing the MT.

Mixed Topology (MT)-Based Cost-Efficient Design Algorithm

For each *(source, destination)* Ethernet connection, do the following:

A- Determine the Connection's Working Path

1) Establish MT for the connection's working path (Working MT) according to MT constructing rules in Figure 2.
2) Stretch the links, i.e., multiply the link weight by a factor α, where $\alpha = \left(\text{Link Rate} \big/ \text{Connection Granularity} \right)$. This will set higher weights on links with high line rates, and hence increase the probability of picking a path with minimum cost. Since they incur no additional cost, MT's virtual (established Etherpaths) link weights are set to zero (to enable grooming).
3) Generate K minimum weight paths (KMWP) on the established MT. If the set contains zero paths, block the connection.
4) For each path in KMWP, calculate cost which equals the number of transceivers of each rate times transceiver cost for that rate.
5) Pick the path with the minimum cost, route the connection, and update the network resources.

B- Determine the Connection's Protection Path

1) Establish MT for the connection's protection path (Protection MT) using MT constructing rules in Figure 2, and perform link stretching.
2) Generate K minimum weight paths (KMWP) over the established MT. If the set contains zero paths, block the connection.
3) For each path in KMWP, find the cost of routing the connection (as determined by Step A-4).
4) Pick the path with the minimum cost, route the connection, and update the network's state.

4. Experimental results

The proposed Algorithms were experimented with on the German network topology shown in Figure 3. Each link is associated with the two numbers (R, D) where R is the link's rate in Gbit/s and D is the link's distance in kilometers. The number of wavelengths on each link is 64. The number of the Ethernet interfaces slots at each node is 256. All links are bidirectional.

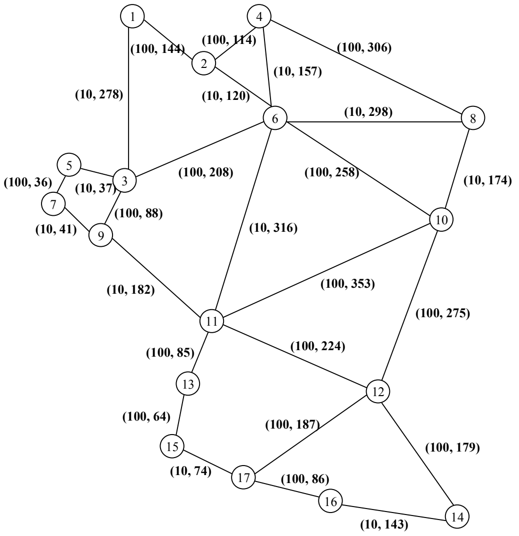

Figure 3: Network topology.

Etherpaths are unidirectional in this current study. The link's rate is the same in both directions. Ethernet connections are 1 Gbit/s, 10 Gbit/s, or 100 Gbit/s. Traffic is asymmetric among different (source, destination) pairs. The traffic matrix is not included because of space limitations. Etherpaths running at 10 Gbit/s and 100 Gbit/s have maximum transmission distances of 3000 km and 500 km respectively, after which regeneration is required. Ethernet interface rates are 10 Gbit/s and 100 Gbit/s with relative costs of 1 and 10 respectively (in this study). Network cost is equal the number of Ethernet interfaces of each rate times the Ethernet interface's cost for that rate. This includes Ethernet interfaces used for setting up *working* Etherpaths and *protection* Etherpaths. Our solution method K

Minimum Weight Paths with Link Stretching (KMWP/s) is compared against the following design algorithms: K Minimum Weight Paths without Link Stretching (KMWP/ns), and K Shortest Paths (KSP). KMWP/s and KMWP/ns construct the MT for generating the minimum-cost paths, while KSP generates the shortest paths without constructing the MT or performing link stretching.

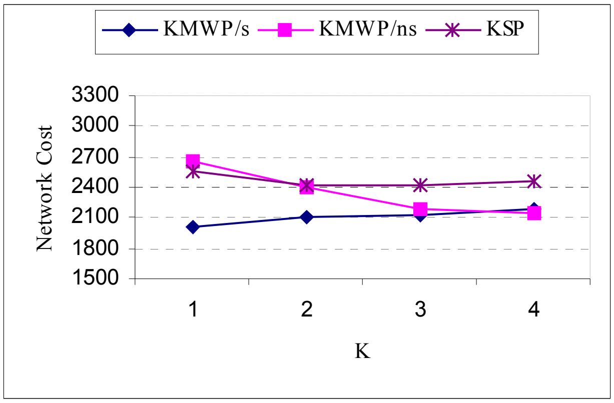

Figure 4: Network cost.

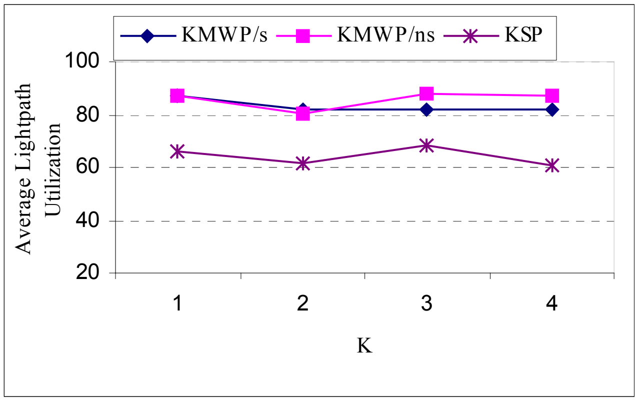

Figure 5: Average utilization of 100 Gbit/s Etherpaths.

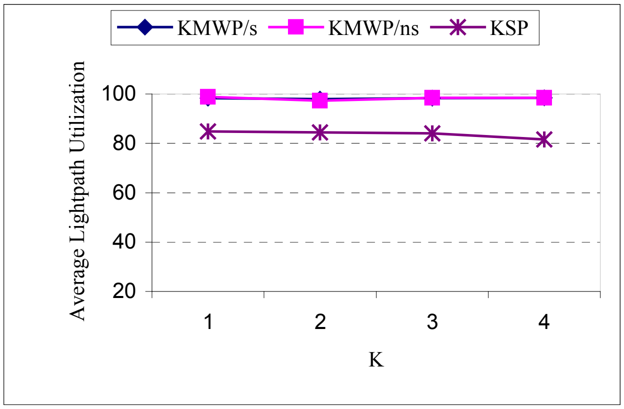

Figure 6: Average utilization of 10 Gbit/s Etherpaths.

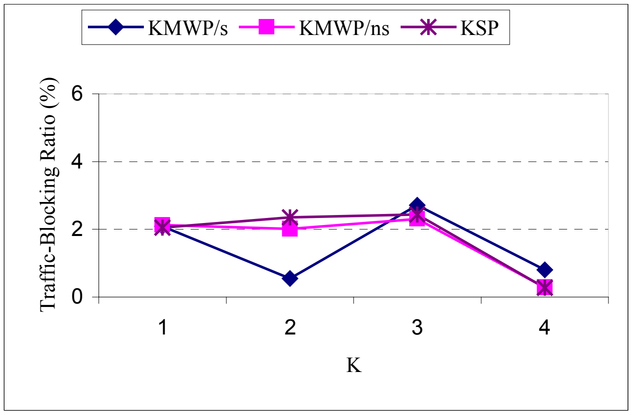

Figure 7: Percentage of blocked traffic.

Figure 4 shows the performance of the three cases with respect to (relative) network cost. KMWP/s shows the minimum CapEx with a savings of 12-25% over other schemes. The minimum cost occurs for KMWP/s at K = 1 because KMWP/s does not need to investigate many paths, making it efficient in terms of computation time. Without link stretching, KMWP/ns's performance degrades but is still better than KSP and the network cost improves as K increases. KMWP/ns needs to test three minimum weight paths (K = 3) to perform close to KMWP/s (see Fig. 4). This shows the efficiency of link stretching, which makes the algorithm aware of the mixed line rates in the network and it will tend to choose paths which reduced cost. For KSP, the network cost is more than KMWP/s for all values of K. In addition, KSP's performance improves slightly from K = 1 to K = 2.

Figures 5 and 6 show the average Etherpath utilization for 100 Gbit/s and 10 Gbit/s, respectively. The utilizations are computed as the overall capacity used for all *working* Etherpaths running at a certain rate divided by the total capacity of all *working* Etherpaths running at that rate. KMWP/s and KMWP/ns perform very similarly, with an average utilization of around 85% and 97%. These numbers reduce to around 60% and 85% for KSP. Figure 7 shows the traffic-blocking ratio (TBR) defined as the amount of traffic blocked divided by the overall amount of traffic offered. TBR shows that for almost the same amount of traffic satisfied, the MT-based algorithms (KMWP/s and KMWP/ns) perform better than the other techniques, resulting in higher CapEx savings and resource usage efficiency.

Two more observations are left: (1) in some cases, as K increases, the network cost increases (or utilization may decrease). Note that this is a greedy approach, and certain path-selection decision may affect the subsequent connections' selected paths. (2) Even though the network cost using KMWP/s is lower than KMWP/ns, the average Etherpath utilization in both cases is very close. Note that the utilization is computed using the working Etherpaths only, where the network cost is based on both working and protection Etherpaths. Also, since no link stretching is performed in KMWP/ns, the probability of selecting a path running at 100 Gbit/s is higher in KMWP/ns than in KMWP/s. Hence, in a 100 Gbit/s Etherpath we expect more Ethernet connections to coexist, so the number of connections sharing the same risk will reduce the number of protection Etherpaths that can be shared for protection. As a result, we need to use more resources for protecting the connections, which will increase the network cost.

Conclusion

We studied the design of reliable and cost-efficient carrier-grade Ethernet using multiple line rates and signal transmission constraints. Our proposed solution algorithms based on Mixed Topology concept and link stretching showed significantly improved performance.

References:

[1] A. Kirstädter, C. Gruber, J. Riedl, T. Bauschert, "Carrier-Grade Ethernet for Packet Core Networks," *Proc., SPIE*, vol. 6354, 2006.
[2] B. Mukherjee, *Optical WDM Networks*, Springer, 2006.
[3] C. Ou, K. Zhu, B. Mukherjee, "Traffic Grooming for Survivable WDM Networks – Shared Protection," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 9, pp. 1367- 1383, Nov. 2003.
[4] K. Zhu and B. Mukherjee, "Traffic Grooming in an Optical WDM Mesh Network," *IEEE JSAC*, vol. 20, no. 1, Jan. 2002.