Optical Network Design for a Multiline-Rate Carrier-Grade Ethernet Under Transmission-Range Constraints

Marwan Batayneh, *Student Member, IEEE*, Dominic A. Schupke, Marco Hoffmann, Andreas Kirstädter, and Biswanath Mukherjee, *Fellow, IEEE*

***Abstract*—Ethernet is a success story in local area networks (LAN). Efforts for extending its boundaries beyond LAN to the carriers' backbone networks are in progress. We study the problem of designing reliable and cost-efficient high-rate (100 Gbit/s) carrier-grade Ethernet in a multiline-rate optical network under signal transmission-range constraints. Reliability is achieved using shared-path protection at the connection level (Ethernet tunnel in this study). We construct an auxiliary graph, called mixed topology (MT), using which it is possible to: 1) identify traffic grooming possibilities; 2) select a path which requires the minimum amount of 3R regeneration; and 3) effectively choose the data rate of the channel to be established. Our algorithms, tested on the 17-node German network, resulted in lower network cost and higher resource utilization compared with other schemes.**

***Index Terms*—3R regeneration, carrier grade, Ethernet, Etherpath, link stretching, mixed topology (MT), multiline rate, signal transmission range.**

I. INTRODUCTION

BEING the dominating LAN technology, around 90% of Internet traffic is generated by end systems with Ethernet interfaces. Since Ethernet is an inexpensive, flexible (plug and play), and widely used technology, it is a strong candidate to be the transport technology of future networks. Hence, efforts for extending Ethernet usage beyond LANs to metropolitan area networks (MAN) and wide area networks (WAN) are in progress.

Traditionally, carrier Ethernet has been transported over other technologies, e.g., SONET/SDH. To achieve this, Ethernet interfaces are connected to SONET/SDH equipment which in turn transport the Ethernet frames in a time-division multiplexing (TDM) fashion. In addition to the increased capital expenditure (CapEx) due to using intermediate layer (SONET/SDH), this operation of Ethernet is also limited by the capabilities and functionalities of the SONET/SDH layer. As transport technologies in general, Ethernet and SONET/SDH show several differences. These can be summarized as follows.

— Line rates of SONET/SDH are 2.5/10/40 Gbit/s and lower. This is clearly a limitation of the envisioned high-rate (100 Gbit/s) Ethernet. Hence, to support 100 Gbit/s Ethernet over SONET/SDH, multiplexing of a 100 Gbit/s signal is required which will result in even higher cost. In addition, the rates of SONET/SDH do not match those of Ethernet (1/10 Gbit/s). In this case, sophisticated mapping schemes are required to transport Ethernet signals over SONET/SDH.

— SONET/SDH is a TDM technology and is originally intended (and optimized) for voice traffic. While traffic growth is dominated by data traffic, Ethernet transport can be used for both data and voice and is a packet-based technology.

— Traditional SONET/SDH uses mainly $1 + 1$ and ring protection. Since Ethernet is intended to transport heterogeneous traffic, it must be able to meet the heterogeneous reliability requirements of the traffic. Hence, as opposed to the rigid fixed protection scheme in SONET/SDH, more flexible and cost-effective schemes can be used, e.g., shared-path protection.

The future mode of operation is to carry native Ethernet frames directly over wavelength-division multiplexing (WDM) optical backbone networks (Ethernet-over-WDM). Thus, several layers of other technologies can be eliminated, and significant savings in CapEx, and operational expenditure (OpEx) can be achieved.

In a backbone network, Ethernet must provide carrier-grade service, i.e., it must provide high level of resilience, flexibility, and manageability. At the same time, it must be a cost-efficient transport technology for the network operator to deploy. Traditionally, Ethernet forwarding is based on the spanning tree protocol (STP). The STP constructs a spanning tree from the root node to every other node in the network, which ensures that any path from the root to any other node is the shortest. But it does not ensure that the path between any other node pair is the shortest. Hence, STP-based approaches used in a backbone network will result in poor resource efficiency. Alternatively, Ethernet can be set up as a connection-oriented service with tunnels carrying Ethernet frames, or simply Ethernet tunnels (ET). Three forwarding technologies for such tunnels are being explored: virtual-LAN crossconnect (VLAN-XC), provider backbone bridge with traffic engineering (PBB-TE), and transport multiprotocol label switching (T-MPLS) [5].

Studies [5], [8] have found that CapEx and OpEx savings can be maximized by running Ethernet over WDM channels (e.g.,

Manuscript received June 28, 2007; revised October 1, 2007.

M. Batayneh is with the Department of Electrical and Computer Engineering, University of California, Davis, CA 95616 USA (e-mail: mmbatayneh@ucdavis.edu).

D. A. Schupke, M. Hoffmann, and A. Kirstädter are with the Nokia Siemens Networks GmbH and Co. KG, RTP NT NCT, D-81730 Munich, Germany (e-mail: dominic.schupke@nsn.com; marco.hoffmann@nsn.com; andreas.kirstaedter@nsn.com).

B. Mukherjee is with the Department of Computer Science, University of California, Davis, CA 95616 USA (e-mail: mukherje@cs.ucdavis.edu).

Digital Object Identifier 10.1109/JLT.2007.913027

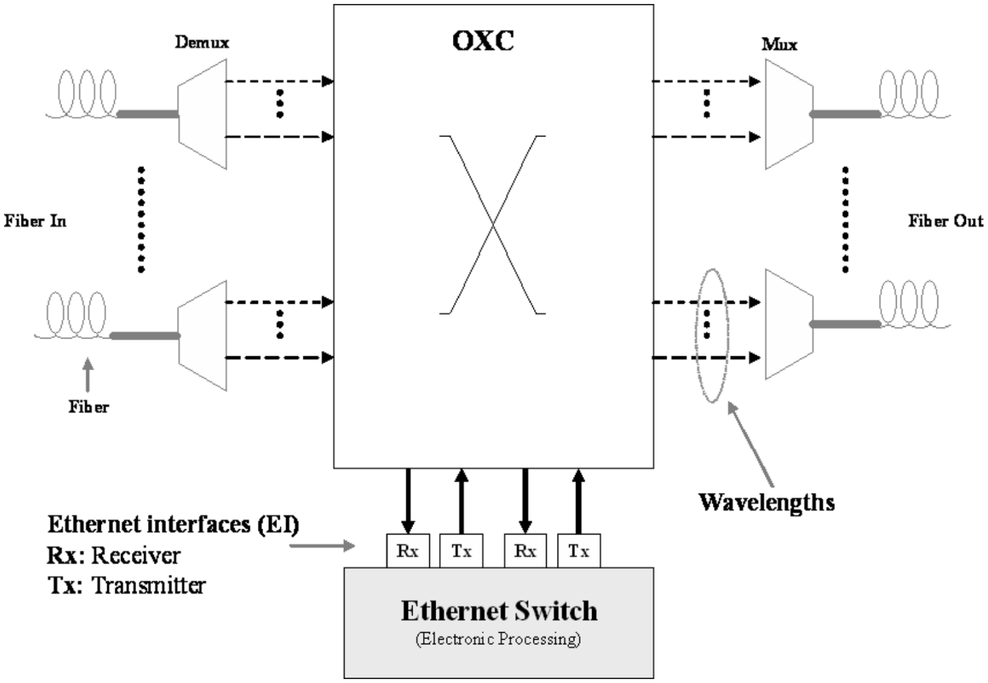


Fig. 1. Node architecture.

lightpaths) at high rates (up to 100 Gbit/s). Since the lightpath rate can be high (100 Gbit/s) and it may travel long distances (for backbone links), the effect of the physical impairments along a path may become prominent and the signal's quality may degrade. Hence, 3R (reamplification, reshaping, and retiming) signal regeneration may be needed [10], [11]. In our study, we use the term Etherpath (EP) to denote a lightpath carrying Ethernet tunnels (ETs) which in turn carry Ethernet frames. These Etherpaths are initiated and terminated by Ethernet interfaces and are solely designed for carrying Ethernet traffic.

If we consider the applications to be transported over future networks, we notice the variation in bandwidth (BW) requirements among them. For instance, voice applications require less BW than video applications, which also require less BW than bandwidth-hungry applications like heavy scientific computations. As a result, the network must be flexible enough to provision these applications effectively. To achieve this, a heterogonous link-capacity distribution is required (multiline-rate network). The term heterogeneous in our current study implies that all wavelengths on a link operate at the same rate, but different links may operate at different rates. This gives the network operator more flexibility in choosing the connectivity (e.g., lightpath) rate according to the available links' rates in the network. In addition, the network operator may also want to operate links with different rates based on their location in the network. For instance, the network operator may use lower rate for links closer to the network periphery, and higher rate for links closer to the network core, since more traffic is expected on the central links. The work in [16] emphasizes this fact by studying the backbone network design with optical crossconnects (OXCs) with different grooming-granularity capabilities (heterogeneous OXCs). The study showed that a network with heterogeneous OXCs is more cost-efficient and flexible than a network with homogenous OXCs.

The problem addressed in this paper is a special case of survivable traffic grooming (STG) [7]. In addition, our problem considers the multiline-rate nature of the network. We also consider the signal's maximum transmission range constraint, which depends on the signal bit rate (as the signal's rate increases, its maximum transmission range decreases). Note that, even though the motivation behind this study is carrier-grade Ethernet, the general concepts and solution methods discussed here also apply to connections carrying traffic other than Ethernet. In the literature as well as throughout the paper, the terms working path and primary path may be used interchangeably to indicate the same thing, i.e., a path that carries the traffic during the network's normal operation. In addition, the terms protection path and backup path may be used interchangeably, i.e., a path that is used to carry the traffic when a single link fails. Likewise, the terms 3R regeneration and regeneration may used interchangeably.

Section II explains the node architecture and 3R regeneration. Section III formally states the problem. Section IV presents the design algorithms. Section V discusses the experimental results. Section VI concludes the study.

II. ETHERNET-OVER-WDM NODE ARCHITECTURE AND 3R REGENERATION

A. Node Architecture

Fig. 1 shows the Ethernet-over-WDM node architecture. It has two components: 1) OXC, which performs switching at the Etherpath level and 2) Ethernet switch (ES). The ES initiates and terminates Etherpaths using Ethernet interfaces (EI), namely Ethernet transmitter and Ethernet receiver. ES performs electronic functions of grooming Ethernet connections onto the Etherpaths, and regeneration. OXC must have enough I/O ports (linecards) to support fiber connections with neighboring nodes as well as for local add/drop with the ES.

B. 3R Regeneration

To perform in-node electronic 3R regeneration (used here), a number of optical-electronic-optical (OEO) transponders (Ethernet interfaces in our case) must be allocated. Now, if an Etherpath requires regeneration, it must be directed from OXC to ES (see Fig. 1), get regenerated, and then sent back to OXC and output fiber. A node with full OEO regeneration capability is called opaque, a node with partial OEO regeneration capability is called hybrid or translucent, while a node without any OEO regeneration capability is called transparent. So, in this paper, we have the hybrid case.

III. PROBLEM STATEMENT

The cost-efficient and reliable carrier-grade Ethernet network design problem is formulated as follows.

- *Given*:
  - —Network topology represented by graph $G = (V, E)$ with set of bidirectional edges E representing links and set of vertices V representing nodes.
  - —Traffic demand matrix composed of Ethernet tunnels with different bandwidth requirements. Ethernet tunnels are unidirectional.
  - —All nodes have regeneration and traffic grooming capabilities.
  - —The network can have mixed line rates.
- *Transmission Constraints*:
  - —Number of wavelengths on a link and wavelength capacity.
  - —Number of Ethernet interfaces at a node.
  - —Maximum transmission range of signals (data rate dependent), i.e., maximum distance an Etherpath can travel before regeneration is required.
  - —Wavelength-continuity constraint is assumed for each Etherpath.
- *Need to*:
  - —Route all Ethernet tunnels.
  - —Protect all Ethernet connections. Shared-path protection-at-connection (PAC) level [7] is used to protect the Ethernet tunnels against single link failures.
- *Objectives*:
  - —Provision all Ethernet tunnels.
  - —Reduce the solution's CapEx: determined by the number and rates of Ethernet interfaces used.

IV. DESIGN ALGORITHMS

In order to design an efficient algorithm, the following issues must be addressed:

1) It is possible that a shorter path, where edge weight is the link's length, requires more amount of 3R regeneration. Before we discuss the example in Fig. 2, Table I illustrates the symbols used throughout the paper. In the example in Fig. 2, maximum (un-regenerated) transmission range is 500 km. Note that the shorter path (900 km) required more regeneration than the longer one (950 km). (The number on a link is its length in km.)

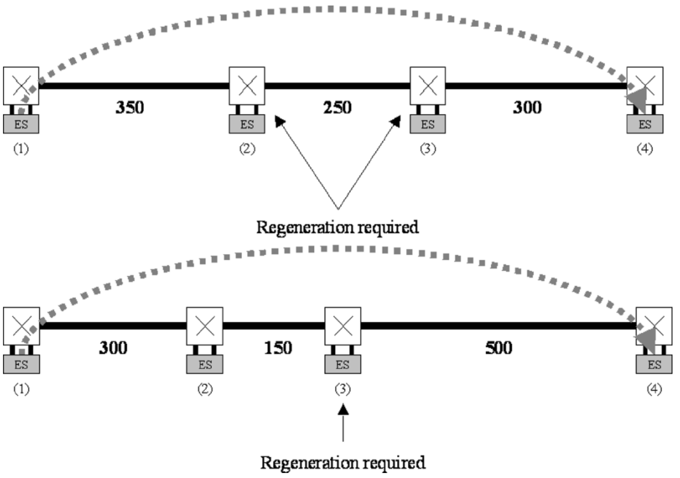


Fig. 2. 3R regeneration example.

TABLE I
SYMBOLS AND THEIR ILLUSTRATIONS

Symbol	Meaning
☒	Optical Cross-connect (OXC)
ES	Ethernet Switch (ES)
━━━	100 Gbit/s Fiber
───	10 Gbit/s Fiber
┅┅▶	100 Gbit/s Etherpath
──→	Ethernet connection

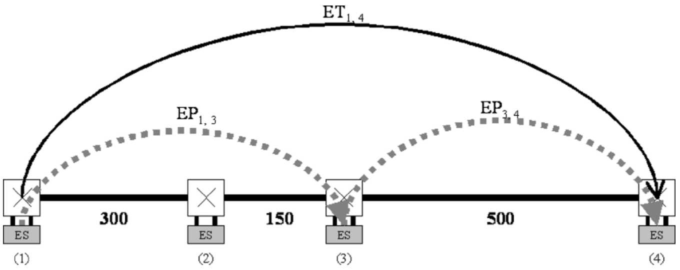


Fig. 3. EP segmentation due to regeneration.

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 (EP) originating from node X and destined to node Z ($\mathrm{EP}_{\mathrm{X-Z}}$) requires regeneration at a third node Y, then since the EP must be terminated in the ES at node Y to perform regeneration, $\mathrm{EP}_{\mathrm{X-Z}}$ is segmented into two Etherpaths, namely, $\mathrm{EP}_{\mathrm{X-Y}}$ and $\mathrm{EP}_{\mathrm{Y-Z}}$. This will create a grooming opportunity into $\mathrm{EP}_{\mathrm{Y-Z}}$ at node Y. It is also possible that the two Etherpaths use different wavelengths.

Fig. 3 shows an example where $\mathrm{ET}_{1,4}$ is to be provisioned. The original EP to carry $\mathrm{ET}_{1,4}$ ($\mathrm{EP}_{1,4}$) had to be dropped at node 3 to perform 3R regeneration (because the distance to the destination is larger than the maximum transmission range). Hence, $\mathrm{EP}_{1,4}$ is segmented into $\mathrm{EP}_{1,3}$ and $\mathrm{EP}_{3,4}$. In this case, it is possible to also groom other traffic onto $\mathrm{EP}_{3,4}$ at node 3 if enough capacity is available.

4) If an Etherpath originating from node X and destined to node Z ($\mathrm{EP}_{\mathrm{X-Z}}$) uses 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 $\mathrm{EP}_{\mathrm{X-Z}}$ must be segmented into two Etherpaths, namely, $\mathrm{EP}_{\mathrm{X-Y}}$ and $\mathrm{EP}_{\mathrm{Y-Z}}$. $\mathrm{EP}_{\mathrm{X-Y}}$ runs at rate R1, and $\mathrm{EP}_{\mathrm{Y-Z}}$ runs at rate R2. Again, this will create a grooming opportunity into $\mathrm{EP}_{\mathrm{Y-Z}}$ at node Y. The two Etherpaths may use different wavelengths.
5) Once an Etherpath is terminated in an Ethernet switch (ES), electronic functions are accessible. For example, grooming can be performed, and wavelength conversion is possible [9]. In addition, regeneration is always performed.

To address the above issues and capture the network state, we introduce the following two tools: (a) mixed topology (MT) and (b) link stretching (LS).

A. MT

The MT is an auxiliary graph. The term mixed is used since the edges of MT might be physical links, virtual links (already established Etherpaths), or both. The MT is constructed on a per-Ethernet-tunnel basis, i.e., for each Ethernet tunnel, the MT is constructed. The MT contains all possible virtual and physical links over which a specific Ethernet tunnel can be routed (i.e., *any path in the MT is a viable route*). MT reflects a partial picture of the network's current state (since not all virtual and physical edges may be established). Using the MT, it is possible to do as follows.:

- Reflect the network's resource usage state by including/excluding the appropriate edges. For instance, if all the wavelengths on a certain link are used, then the link cannot be included in the MT. At the same time, if any of its used wavelengths has some residual capacity, then a *virtual* (unidirectional) edge can be included to reflect this available capacity.
- Implement any design policy by including/excluding the appropriate physical/virtual edges, and assigning weights to these edges. For instance, if we need to reduce the number of high-rate EPs (since they require more regeneration), then high-rate links may be assigned higher weights.
- Reflect the traffic grooming (TG) and wavelength conversion opportunities through EP segmentation.
- Find the ET's protection path. Establishing the MT for finding a protection path must address the resource-sharing constraints. Hence, risk information must be maintained (as well as other resource constraints). For instance, MT for $\mathrm{ET_x}$, which shares the same risk with another tunnel $\mathrm{ET_y}$, must not include the protection EPs of $\mathrm{ET_y}$.

For each ET, the MT is constructed to determine both the primary and the protection paths. To determine the ET's primary path, the MT must include the links and EPs that can carry the ET. Using shared-path protection, each ET is protected against any single link failure. In this case, the MT for determining the ET's protection path includes all the physical links and EPs can then be used for protecting the ET (i.e., all EPs and links that do not violate the resource-sharing constraints).

In our study, we use the following rules to construct the MT for both primary and protection paths.

A) The MT for finding the primary path includes all the physical and virtual links except:
   1. all links running at rates below the ET's rate;
   2. all EPs reserved for protection;
   3. all saturated EPs or EPs with free capacity below the ET's rate;
   4. all saturated links (saturated means all wavelengths are used);
   5. if the end nodes of a physical link or a physical segment (set of adjacent physical links) are connected by a working EP, then this link (or segment) is replaced by the EP (assuming the EP has enough capacity to carry the ET).

B) The MT for finding the protection path includes all the physical and virtual links except:
   1. all links running at rates below the ET's rate;
   2. all physical links on the ET's primary path;
   3. all protection EPs for tunnels sharing the same risk with the current tunnel if the aggregate capacity of the tunnels exceeds each 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. physical link or a physical segment (set of adjacent physical links) with both end nodes connected by a protection EP is replaced by the EP (assuming the rate of the protection EP is higher than the ET's rate and it does not violate the resource-sharing constraints).

Notice that, depending on the network design objectives and the protection scheme used, different rules may be used for constructing both primary and protection MTs.

B. LS

For an ET with a certain bandwidth requirement, it is possible to carry the ET over EPs with different rates. For instance, we can carry a 1-Gbit/s ET over a 10-Gbit/s EP or over a 100-Gbit/s EP. If we establish a 100-Gbit/s EP, we need to use 100 Gbit/s Ethernet interfaces or we may establish a 10-Gbit/s EP where we need to use 10 Gbit/s Ethernet interfaces which are much cheaper than 100 Gbit/s interface.

To address this issue, and to increase the possibility of picking the minimum-cost path, we introduce the concept of LS. Using LS, the link's weight is multiplied (stretched) by factor α that depends on both the link's rate and the ET's rate. Factor α is defined as

$$\alpha = \frac{\text{Link Rate}}{\text{ET Rate}}. \tag{1}$$

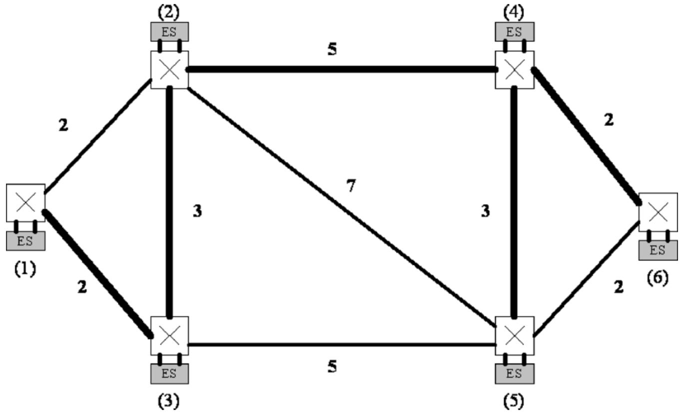


Fig. 4. Example network.

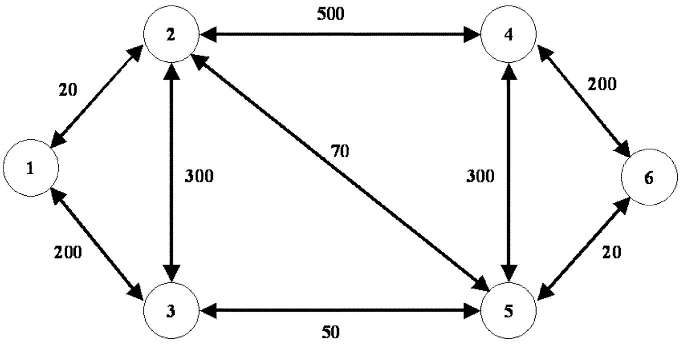


Fig. 5. MT for $\mathrm{ET}_1(2,6,1)$.

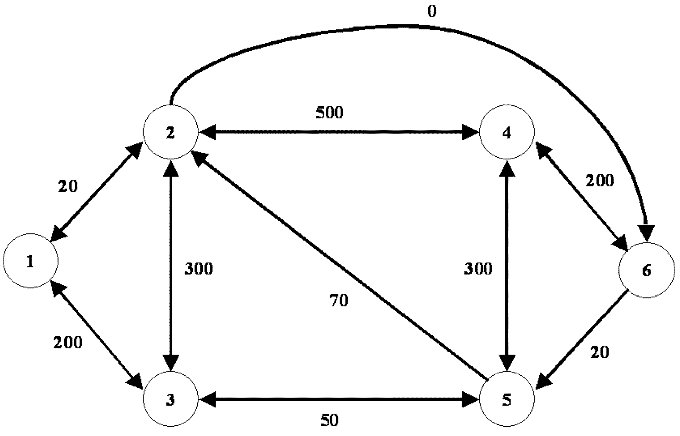


Fig. 6. MT for $\mathrm{ET}_2(1,6,1)$.

Using link stretching, it is possible to do as follows:

- set higher weights on links with high rates. Hence, if we generate a set of shortest paths, the probability of generating shorter (low-rate and cost-efficient) paths will increase.
- reduce the 3R regeneration required. This is because generating low-rate paths means higher transmission range; and, hence, less 3R regeneration.
- enable grooming on established EPs, where weights on the MT's virtual links (already established EPs) are set to zero.
- improve network throughput (the amount of traffic satisfied) by reducing traffic blocking. Using MT, each ET is either groomed onto an available EP, satisfied by establishing new EPs, or both [17]. If new EPs are established, LS will guarantee that the most appropriate EP rate is selected to provision an ET. Hence, LS tries to pack as much traffic into a given amount of capacity as possible. Since packing is known to save resources, it is capable of satisfying more traffic and reducing traffic blocking.

Consider the example in Fig. 4 where the thick links operate at 100 Gbit/s and the thin ones operate at 10 Gbit/s. Link weight is its length in hundreds of kilometers. For illustration purpose, let each link have a single wavelength. Links are bidirectional and Etherpaths are unidirectional. Etherpaths running at 10 and 100 Gbit/s have maximum transmission distances of 3000 and 500 km, respectively, after which regeneration is required. The following ETs are to be satisfied in the order shown: $ET_1(2,6,1)$, $ET_2(1,6,1)$, $ET_3(2,5,100)$, and $ET_4(3,6,1)$. The general format of the ET, ET (s, d, BW) is explained as follows: s and d are the source and destination nodes, respectively, and BW is the ET's bandwidth requirement in Gbit/s.

If we generate the shortest path based on the original link weights (Fig. 4), then the shortest path for ET_1 is $\{2-4-6\}$. If ET_1 follows this path, it will require four 100 Gbit/s EIs, two for initiating and terminating the EP at nodes 2 and 6, and two for performing 3R regeneration at node 4. Using LS, the new weight assignment will set higher weights on the 100 Gbit/s links, and the shortest path will be $\{2-5-6\}$, which requires two 10 Gbit/s EIs. The MT that corresponds to ET_1 is shown in Fig. 5. All the physical links (bidirectional edges) are included since they are capable of carrying ET_1 and have enough resources. Notice the link weight assignment based on LS.

The MT for ET_2 is shown in Fig. 6. Notice that links (2,5) and (5,6) are included in a single direction (shown in Fig. 6). This is because one wavelength channel in the other direction is already used for setting up the EP carrying ET_1 ($\mathrm{EP}_{2,\,6}$). In addition, a new (virtual) edge connecting nodes 2 and 6 is added. This edge reflects the free capacity (9 Gbit/s) available in $\mathrm{EP}_{2,\,6}$. In this case, the routing for ET_2 is $\{1-2-6\}$ using the link 1–2 and the $\mathrm{EP}_{2,\,6}$ which directly connects nodes 2 and 6. The cost of this path is two 10 Gbit/s EIs. It is possible to use the path $\{1-2-5-6\}$ with the same cost of two 10 Gbit/s EIs. But in this case, a single wavelength channel must be found along the three links of path $\{1-2-5-6\}$. Using path $\{1-2-6\}$, only a single wavelength channels on link 1–2 is required. Hence, using this path, it is more possible that ET_2 gets successfully provisioned, and at the same time, more wavelength channels have been saved for use by other tunnels.

For ET_3, the MT is shown Fig. 7. All links running below 100 Gbit/s are not included. In addition, $\mathrm{EP}_{1,\,2}$ (created to support ET_2) and $\mathrm{EP}_{2,\,6}$ are not included as they do not have enough capacity to carry ET_3. Using this MT, there is only a single path from node 2 to node 5, i.e., $\{2-4-5\}$. Since the path length is larger than the maximum transmission range (500 km), 3R regeneration is required. Hence, the EP that carries ET_3 (i.e., $\mathrm{EP}_{2,\,5}$) must be segmented at node 4 (due to 3R regeneration) into two EPs, namely, $\mathrm{EP}_{2,\,4}$ and $\mathrm{EP}_{4,\,5}$. The cost of this path is four 100 Gbit/s EIs.

Finally, the MT of ET_4 is shown Fig. 8. Notice that $\mathrm{EP}_{2,\,4}$ and $\mathrm{EP}_{4,\,5}$ are not included as they do not have free capacity.

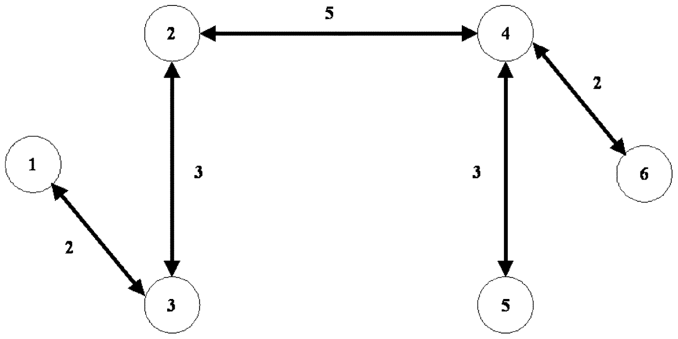


Fig. 7. MT for ET3 (2, 5, 100).

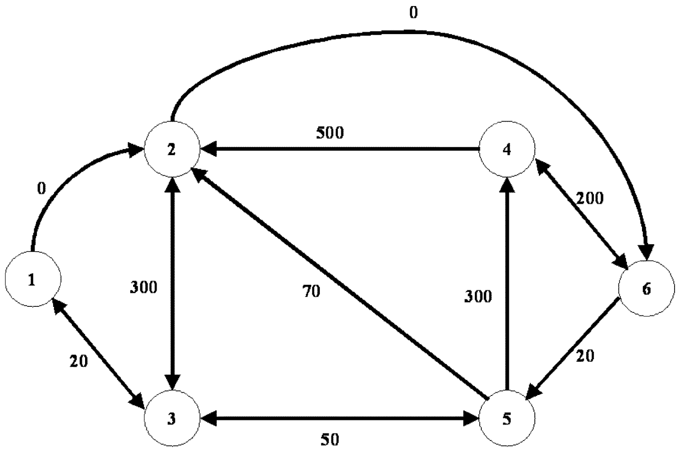


Fig. 8. MT for $\mathrm{ET}_4(3, 6, 1)$.

The route of this tunnels is {3 – 1 – 2 – 6} with a cost of two 10 Gbit/s EIs. In addition, only a single wavelength channel is required to provision ET_4.

If we need to construct the MT for a protection path, the same treatment as in the previews examples applies, with the exception that resource-sharing constraints will also determine how the physical/virtual edges are included (see rules for constructing the protection MT).

C. Solution Algorithms

Now, based on the MT construction rules and the concept of link stretching, we propose the following algorithm: MT-based algorithm for reliable and cost-efficient carrier-grade Ethernet design using PAC [7] and link stretching, or simply MT/s. The algorithm is explained as follows.

For each Ethernet tunnel, do the following:

A) *Determine the ET's working path*:

Step 1) Establish MT for ET's working path (Working MT).

Step 2) Stretch the links to reflect the link's cost (rate dependence) and 3R regeneration requirement.

Step 3) Generate a set of K minimum-weight paths (KMWP) [14] over the established MT. If the set contains zero paths, block ET.

Step 4) For each path in KMWP, calculate cost of using that path. Path cost is equal to the number of Ethernet interfaces of certain rate times the Ethernet interface's price for that rate.

Step 5) Pick path with minimum cost, route the tunnel, and update usage information on the network resources.

B) *Determine the ET's protection path*:

Step 1) Establish MT for ET's protection path (Protection MT) and perform link stretching.

Step 2) Generate set of KMWP. If set contains zero paths, block ET.

Step 3) For each path in KMWP, find cost of routing the ET. (Cost is determined as in Step A-4.)

Step 4) Pick path with minimum cost, reserve resources over that path, and update the network's state.

A set of paths is generated for two reasons: 1) to find a path that requires less regeneration (since shorter path may sometimes require more regeneration); and 2) if a single wavelength is not found along the shortest path, the next shortest path is examined and so on. Hence, this will increase the possibility of provisioning the ET. The K minimum-weight paths used in the algorithm are not necessarily link-disjoint.

In this algorithm, routing and wavelength assignment (RWA [1], [4], [6], [15]) are combined, i.e., as soon as a path is found, the number (and rate) of EPs to be established are determined; and for each EP a single (wavelength-continuous) path is searched along its route. If such a path is not found, the Ethernet tunnel is blocked.

V. EXPERIMENTAL RESULTS

The proposed solution methods were experimented with on the 17-node German network topology shown in Fig. 9. The number of wavelengths on each 100 Gbit/s link is 16, and the number of wavelengths on each 10 Gbit/s link is 128. The link rates are shown in the figure. The number of Ethernet interface slots at each node is 128. Each interface slot is assumed to operate in a range of rates (10–100 Gbit/s), which enables the slot to support either 10 Gbit/s EI or 100 Gbit/s EI. Links are bidirectional. Etherpaths are unidirectional. A link has a single rate, same in both directions. Ethernet tunnels are at 1 Gbit/s, 10 Gbit/s, or 100 Gbit/s.

The network design parameters (cost, traffic-blocking ratio, and average Etherpath utilization) are plotted against the traffic. The methodology for generating the traffic is as follows. The traffic matrix in Fig. 10 is used as the base matrix. The traffic matrix is scaled (multiplied) by a factor from the set {0.25, 0.5, 0.75, 1.0, 2.0, and 3.0}. The *integer* outcome of this multiplication is then taken. Hence, it is possible that some entries become zero. These scalars correspond to the following aggregate traffic ($= \text{sum of all traffic entries in the resulting traffic matrix}$): {460, 1040, 1458, 2212, 4424, and 6636} Gbit/s. Etherpaths running at 10 and 100 Gbit/s have maximum transmission distances of 3000 and 500 km, respectively, after which regeneration is required.

Based on current industry trends, Ethernet interface rates of 10 Gbit/s and 100 Gbit/s have relative costs of 1 and 5, respectively. Network cost is equal to the number of Ethernet interfaces of each rate times the Ethernet interface's cost for that rate.

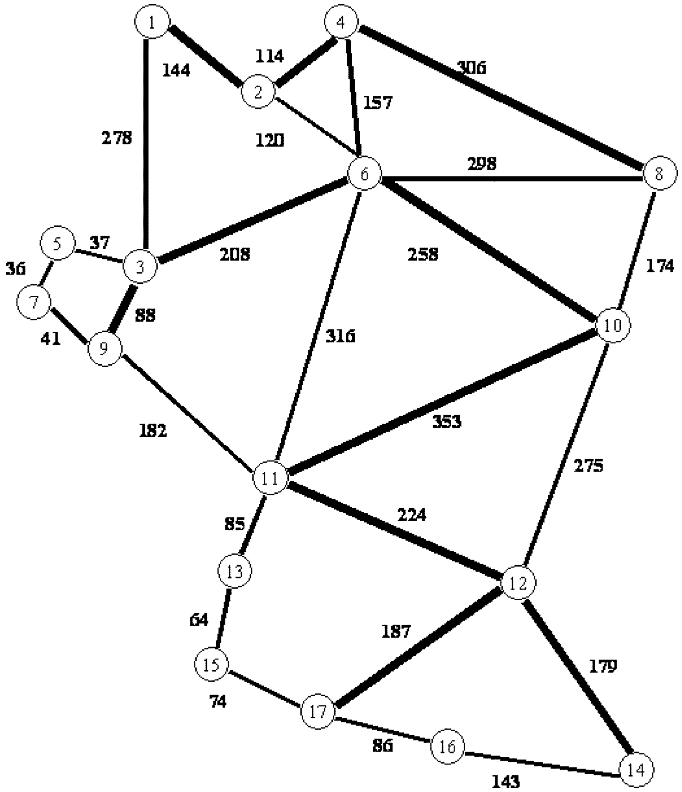


Fig. 9. German network topology (thick line = 100 Gbit/s, thin line = 10 Gbit/s, and link metric = link length in km).

This includes Ethernet interfaces used for setting up *working* Etherpaths and *protection* Etherpaths.

The design algorithms compared against each other are as follows:

- *MT/s:* Algorithm is based on constructing the MT and performing link stretching;
- *MT/ns:* Algorithm is based on constructing the MT but no link stretching is used;
- *SP:* Algorithm is based on a link's physical distance to generate the shortest path, i.e., neither MT nor link stretching is used for finding the working path. To find the protection path, protection MT is constructed to avoid violating the resource-sharing constraints (stretching is not used).

Figs. 11 and 12 show the performance of the three cases in terms of their impact on the network cost. The cost shown in the figures is normalized cost, i.e., all cost values are divided by a base value. This value is the lowest cost achieved by all the algorithms at any traffic value, which is 426 (cost of the Ethernet interfaces used) and is achieved by MT/s at traffic scalar 0.25 (Fig. 11) (normalized cost = 1). It is also used as the base value for the cost values in Fig. 12. For fairness, the comparison is tied to the traffic-blocking ratio (TBR) shown in Fig. 13. TBR is determined as follows:

$$\mathrm{TBR} = \frac{\text{traffic blocked}}{\text{total traffic}}. \qquad (2)$$

TBR is incorporated in the cost calculation since this is a network design problem. In network design, all traffic must be satisfied, and a higher TBR may lead to higher cost (in order to reflect the cost of the blocked traffic). Hence, we incorporate the amount of successfully routed traffic in computing the algorithm's cost efficiency. In this case, the cost is computed as follows:

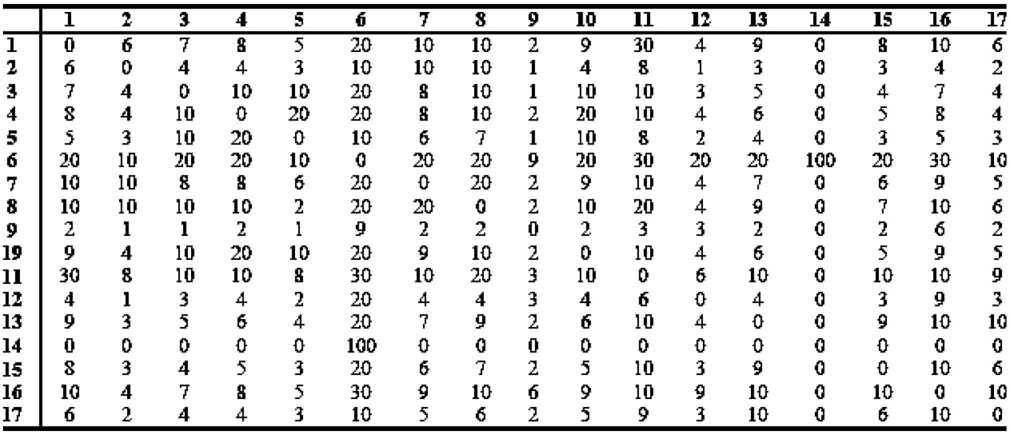
$$\mathrm{Cost} = \frac{\text{normalized cost}}{(1 - \mathrm{TBR})}. \qquad (3)$$

The general observation is that MT/s results in lower cost for all traffic values. In addition, the MT-based algorithms (MT/s and MT/ns) result in lower cost than the non-MT-based algorithm (SP) at all the traffic scalar values. Traffic is divided into two regions: 1) low-traffic region (traffic multiplier less than 1.0); and 2) high-traffic region (traffic multiplier 1.0 and above).

In low-traffic region (Fig. 11), MT/s achieves low network cost since it efficiently selects the suitable EP rates and performs grooming as well. Notice that TBR for the three algorithms is close (around zero) in this region, even though MT/s achieved the lowest cost which shows its efficiency in using the resources (Ethernet interfaces and wavelength channels). In high-traffic region (Fig. 12), MT/s also achieves lower cost than MT/ns and SP. This region is characterized by high TBR due to the resource and wavelength-continuity constraints. Hence, we notice that MT/s is resource efficient at various loads. In fact, its efficient performance at low-traffic loads plays an important role in making it also efficient at high-traffic loads. This is because, at low-traffic loads, resources are effectively used and more excess capacity [2] can be spared for later use at high-traffic loads.

If we consider TBR (Fig. 13), we notice the following:

1) There is some increase in TBR for MT/s versus MT/ns (especially in the high-traffic region), which can be explained as follows. Using link stretching, the algorithm is biased toward low-rate links, and Ethernet tunnels tend to use low-rate wavelength channels. In this case, the low-rate channels are consumed quickly. Due to wavelength-continuity constraint, it is possible that some tunnels may not find a single wavelength channel available on all the low-rate links on the minimum-weight paths generated by the algorithm, especially the tunnels that the algorithm tries to provision after large amount of traffic has been provisioned and few wavelength channels are available. Furthermore, this is more likely to occur at higher loads.
2) Non-MT-based algorithm (SP) results in high TBR compared to the MT-based algorithms. This is explained as follows. The SP algorithm depends only on the actual length of the links, and it does not efficiently perform grooming since it lacks the intelligence provided to the MT-based algorithms by the construction of MT. At higher loads, the wavelength-continuity constraint will be a limiting factor on the success of satisfying the traffic: it is hard to find the same wavelength available on all links of a path. In addition, since SP does not perform grooming efficiently, the EP may need to pass through more physical links, and the average hop distance (physical in this case) of the EP may become longer. Hence, the wavelength-continuity constraint will become even more constraining.



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	0	6	7	8	5	20	10	10	2	9	30	4	9	0	8	10	6
2	6	0	4	4	3	10	10	10	1	4	8	1	3	0	3	4	2
3	7	4	0	10	10	20	8	10	1	10	10	3	5	0	4	7	4
4	8	4	10	0	20	20	8	10	2	20	10	4	6	0	5	8	4
5	5	3	10	20	0	10	6	7	1	10	8	2	4	0	3	5	3
6	20	10	20	20	10	0	20	20	9	20	30	20	20	100	20	30	10
7	10	10	8	8	6	20	0	20	2	9	10	4	7	0	6	9	5
8	10	10	10	10	2	20	20	0	2	10	20	4	9	0	7	10	6
9	2	1	1	2	1	9	2	2	0	2	3	3	2	0	2	6	2
19	9	4	10	20	10	20	9	10	2	0	10	4	6	0	5	9	5
11	30	8	10	10	8	30	10	20	3	10	0	6	10	0	10	10	9
12	4	1	3	4	2	20	4	4	3	4	6	0	4	0	3	9	3
13	9	3	5	6	4	20	7	9	2	6	10	4	0	0	9	10	10
14	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0
15	8	3	4	5	3	20	6	7	2	5	10	3	9	0	0	10	6
16	10	4	7	8	5	30	9	10	6	9	10	9	10	0	10	0	10
17	6	2	4	4	3	10	5	6	2	5	9	3	10	0	6	10	0

Fig. 10. Traffic matrix (Gbit/s).

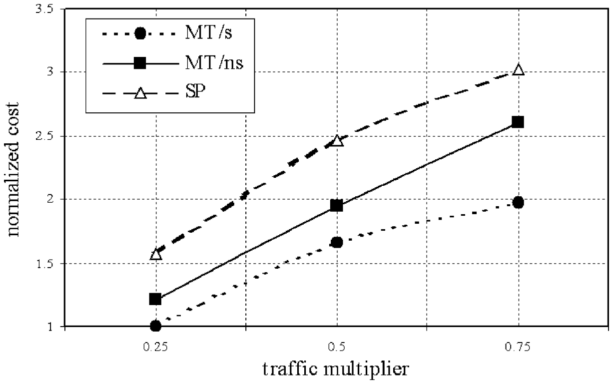


Fig. 11. Network cost in low-traffic region.

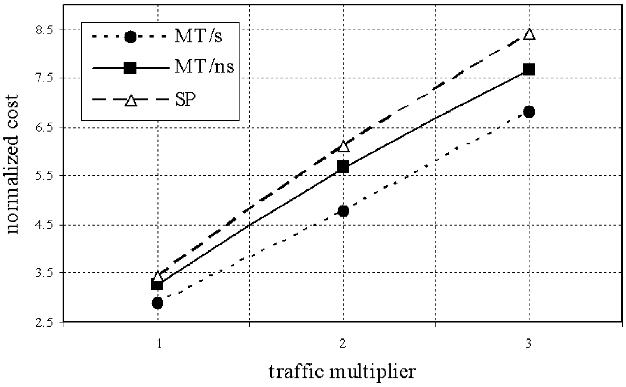


Fig. 12. Network cost in high-traffic region.

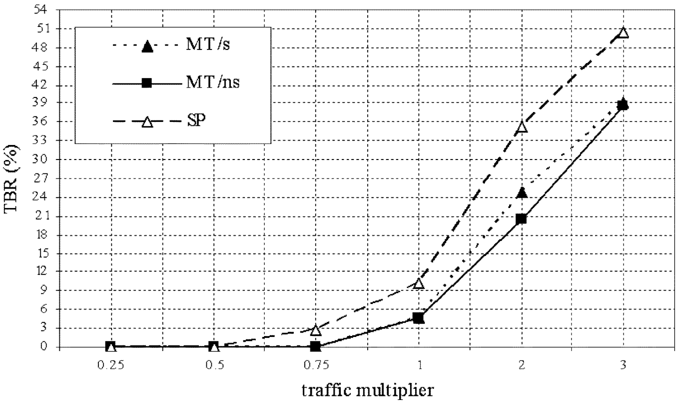


Fig. 13. Traffic-blocking ratio (TBR).

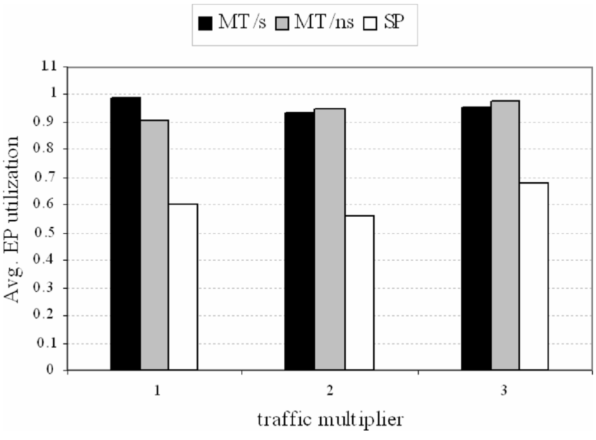


Fig. 14. Average Etherpath utilization.

We comment on one more observation regarding the average EP utilization (Fig. 14). In general, MT-based algorithms (MT/s and MT/ns) achieved much higher EP utilization (around 95%) compared to SP algorithm (around 60%), which further shows the resource efficiency of the MT-based algorithms. In addition, even though the network cost using link stretching (MT/s) is lower, the average EP utilizations in both MT/s and MT/ns are very close where we may expect MT/s to achieve higher utilization. This is justified as follows.

1) Since the algorithm sets zero weights on the MT's virtual (Etherpath) edges, the Ethernet tunnels will be strongly forced to follow these EPs. For example, the cost of a path with one edge with weight W1 and one virtual edge (weight $= 0$) is larger than the cost of a path with one edge with weight $W2 < W1$ and N virtual edges. Hence, the connection will use the second path, filling all the N EPs (virtual edges), which will increase their utilization. From cost perspective, this path is preferred, but in reality, resources (wavelength channels) are being used which may affect the subsequent connections and increase the network cost. We call this kind of utilization as fake utilization in the sense that many EP have been used unnecessarily. In MT/ns, since more 100 Gbit/s EPs are established, these

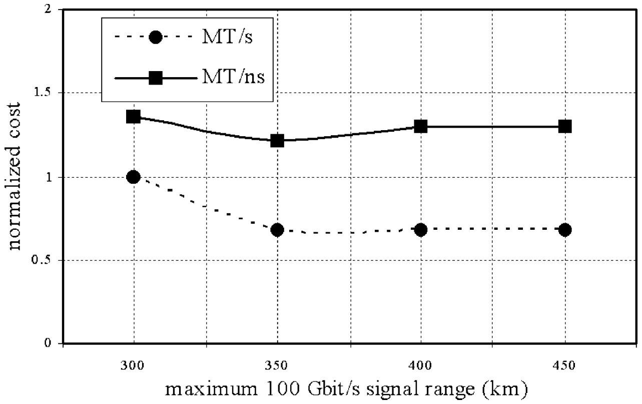


Fig. 15. Network cost versus transmission reach.

EPs will take longer to be filled, and hence MT/ns is more affected by the fake utilization phenomena, which in turn increases the average EP utilization. A good example of such a case is the MT of $\mathrm{ET}_4(3, 6, 1)$ shown in Fig. 8. In this example, ET_4 uses link (3, 1), $\mathrm{EP}_{1,2}$, and $\mathrm{EP}_{2,6}$ instead of link (3, 2) and $\mathrm{EP}_{2,6}$.

2) Since no link stretching is performed in MT/ns, the probability of selecting a path running at 100 Gbit/s is higher. Hence, in a 100 Gbit/s EP, we expect more Ethernet connections to coexist, so the number of connections sharing the same risk will reduce the number of backup EP 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. Note that the utilization is computed using the working EP only, while the network cost is based on both working and protection EP.

Finally, we consider the effect of the maximum transmission reach of the signal on the network's cost. For this part, we assume all source-destination pairs (tunnels) to have the same traffic requirement of 10 Gbit/s. No protection is accounted for in this experiment. The maximum reach of the 10 Gbit/s is assumed to be 3000 km so that no 3R regeneration will be required for the 10 Gbit/s EPs using the network in Fig. 9. The maximum reach of the 100 Gbit/s signal is $\{300, 350, 400, \text{and } 450\}$ kilometers, after which 3R regeneration is required. Network cost is computed as in (3). All cost values are normalized against the cost captured by MT/s at maximum reach value of 300 km (the lowest). Furthermore, only MT/s and MT/ns are considered in this experiment. This is because the experiment examines the effect of LS on reducing 3R regeneration which is the only distinction between MT/s and MT/ns. Fig. 15 shows the network cost sensitivity against the maximum 100 Gbit/s signal transmission range.

Based on Fig. 15, we make the following observations.

1) MT/ns results in higher cost than MT/s. This is mainly because MT/ns does not use LS. Hence, it may generate more 100 Gbit/s EPs and require more 3R regeneration. In addition, since MT/ns may use more 100 Gbit/s EPs, the wavelength-continuity constraint will be a more limiting factor in MT/ns than in MT/s because the 100 Gbit/s links have fewer wavelengths and it may be harder to find the same wavelength along the path.

2) As the transmission reach is increased from 300 to 350 km, the network cost reduces for both algorithms. This is because increased range reduces the amount of 3R regeneration required. As the reach increases to 350 km and beyond, the network cost decrease becomes almost flat, especially for MT/s. This is because: a) the length of the generated path is less than the maximum transmission range, or b) since both MT/s and MT/ns use the MT, they can groom the ETs onto the already-established EPs (virtual edges) which will decrease the length of a newly established EP (if they need to establish ones to satisfy the ETs [17]) and hence, the 3R regeneration will decrease. For MT/s, this behavior is clearer since fewer 100 Gbit/s EPs are established and hence the effect of transmission range is less than the case of MT/ns. We also notice that, for MT/ns algorithm, there is some increase in cost as we go from 350 to 400n km, where we may expect the opposite. This increase is not related to the reach increase, but it is caused by the routing and wavelength selection decisions.

3) The results shown in Fig. 15 are specific to the traffic assumed (10 Gbit/s tunnels) among source-destination nodes and the German network topology. It is possible to get different outcomes using different networks, link-rate distributions, and traffic characteristics. For example, as the number of high-rate (e.g., 100 Gbit/s) ETs increases, the effect of 3R regeneration will increase and the performance of MT/s and MT/ns may be close (because MT/s is forced to use 100 Gbit/s channels). On the other hand, if the low-rate (e.g., 10 Gbit/s) ETs increases, MT/s may perform better than MT/ns.

VI. CONCLUSION

We studied the design of reliable and cost-efficient carrier-grade Ethernet in a multiline-rate network with signal-transmission-range constraints with high rates of operation (100 Gbit/s). These constraints are very practical in a carrier's network. We proposed a graph-based solution, which depends on constructing the MT and the novel concept of *link stretching*. The approach showed improved performance compared to non-MT-based approaches. In addition, even though protection-at-connection (PAC) level is used in this study, protection-at-lightpath (PAL) can also be used. Our study in [3] makes a performance comparison between PAC and PAL for carrier-grade Ethernet.

REFERENCES

[1] D. Banerjee and B. Mukherjee, "Wavelength-routed optical networks: Linear formulation, resource budgeting tradeoffs, and a reconfiguration study," *IEEE/ACM Trans. Netw.*, vol. 8, pp. 598–607, Oct. 2000.

[2] M. Batayneh, S. Rai, S. Sarkar, and B. Mukherjee, "Efficient management of a network's excess capacity: A traffic engineering approach," presented at the Europ. Conf. Opt. Commun. (ECOC), Sep. 2007.

[3] M. Batayneh, D. A. Schupke, M. Hoffmann, A. Kirstaedter, and B. Mukherjee, "Lightpath-level protection versus connection-level protection for carrier-grade Ethernet in a mixed-line-rate telecom network," presented at the IEEE Globecom, Nov. 2007.

[4] R. Dutta and G. N. Rouskas, "A survey of virtual topology design algorithms for wavelength routed optical networks," *Opt. Netw. Mag.*, vol. 1, no. 1, pp. 73–89, Jan. 2000.

[5] A. Kirstädter, C. Gruber, J. Riedl, and T. Bauschert, "Carrier-grade Ethernet for packet core networks," in *Proc. SPIE*, 2006, vol. 6354.

[6] B. Mukherjee, *Optical WDM Networks*. New York: Springer, 2006.

[7] C. Ou, K. Zhu, and B. Mukherjee, "Traffic grooming for survivable WDM networks – Shared protection," *IEEE J. Sel. Areas Commun.*, vol. 21, pp. 1367–1383, Nov. 2003.

[8] A. Schmid-Egger and A. Kirstädter, "Ethernet in core networks: A technical and economical analysis," in *Proc. HPSR'06*, Poznan, Poland, Jun. 2006, pp. 135–140.

[9] D. A. Schupke, M. Jäger, and R. Hülsermann, "Comparison of resilience mechanisms for dynamic services in intelligent optical networks," in *Proc. 4th Int. Workshop on the Design of Reliable Commun. Networks (DRCN)*, Banff, Canada, 2003, pp. 106–113.

[10] J. M. Simmons, "Network design in realistic all-optical backbone networks," *IEEE Commun. Mag.*, vol. 44, no. 11, pp. 88–94, Nov. 2006.

[11] J. M. Simmons, "On determining the optimal optical reach for a long-haul network," *J. Lightw. Technol.*, vol. 23, no. 3, pp. 1039–1048, Mar. 2005.

[12] S. Thiagarajan and A. Somani, "Traffic grooming for survivable WDM mesh networks," *Opt. Netw. Mag.*, vol. 3, no. 3, pp. 88–98, May/Jun. 2002.

[13] J. W. Yao and B. Ramamurthy, "Constrained dynamic traffic grooming in optical WDM mesh networks with link bundled auxiliary graph model," in *Proc. IEEE HPSR'04*, Phoenix, AZ, April 2004, pp. 287–291.

[14] J. Yen, "Finding the K shortest loopless paths in a network," *Manage. Sci.*, vol. 17, no. 11, Jul. 1971.

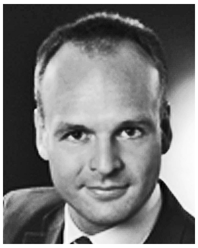
[15] H. Zang, J. P. Jue, and B. Mukherjee, "A review of routing and wavelength assignment approaches for wavelength-routed optical WDM networks," *Opt. Netw. Mag.*, vol. 1, no. 1, pp. 47–60, Jan. 2000.

[16] H. Zhu, K. Zhu, H. Zang, and B. Mukherjee, "Cost-effective WDM backbone network design with OXCs of different bandwidth granularities," *IEEE J. Sel. Areas Commun.*, vol. 21, pp. 1452–1466, Nov. 2003.

[17] K. Zhu and B. Mukherjee, "Traffic grooming in an optical WDM mesh network," *IEEE J. Sel. Areas Commun.*, vol. 20, pp. 122–133, Jan. 2002.

Marwan Batayneh (S'07) received the B.S. degree from Jordan University of Science and Technology, Irbid, and the M.S. degree from the University of California, Davis, all in electrical and computer engineering in 2001 and 2006, respectively.

He is currently a Ph.D. degree candidate with the Department of Electrical and Computer Engineering, University of California, Davis, where he is a graduate student researcher in the Networks Lab. His research focus is on the design and analysis of carrier-grade Ethernet architectures.



Dominic A. Schupke studied electrical engineering at RWTH Aachen University, Germany and at Imperial College London, U.K. He received the Dipl.-Ing. degree from RWTH Aachen in 1998, and the Dr.-Ing. degree (with distinction) from the Munich University of Technology (TUM), Munich, Germany, in 2004.

From 1998 to 2004, he was research and teaching staff member of the Institute of Communication Networks, TUM. At the Institute, he was involved in research projects with both industrial and academic partners, and, in summer 2003, he was visiting researcher at TRLabs, Edmonton, Canada. In February 2004, he joined the Research and Development Department, Siemens, Corporate Technology, Munich. There, he became Senior Research Scientist and Competence Field Manager of the Network Configuration and Planning Group in the Optical Networks and Transmission Department. In October 2006, the department was transferred to Siemens Networks, which then has become part of the joint-venture Nokia Siemens Networks (NSN) in April 2007. His research interests are in the area of transport networks. This includes network architectures and protocols (optical networks, IP/MPLS networks, Ethernet networks, multilayer networks), routing and wavelength assignment, recovery methods, availability analysis, critical infrastructures, network optimization, and network planning. He is the author or coauthor of 10 journal contributions and more than 40 conference papers.

Dr. Schupke has been a Technical Program Committee Member for IEEE conferences ICC and Globecom, and for OFC, DRCN, ONDM, and Chinacom. Since March 2007 he has served as Associate Editor of the *OSA Journal of Optical Networking*. He was also Editor for a Special Issue of the *International Journal of Electronics and Communications* (Elsevier) in 2006. He is a member of IEEE/Comsoc, VDE/ITG, and VDI.

Marco Hoffmann studied computer science at the Munich University of Technology (TUM), Munich, Germany. He recieved the Dipl.-Inform. degree and the Dr. rer. nat. degree, both from TUM, in 2001 and 2005, respectively.

In 2001, he joined the Research and Development Group, GPRS base station Software of Siemens AG, Munich. From the end of 2001 to April 2004, he was a research and teaching staff member with the Institute of Computer Science, TUM. At the Institute, he was the TUM project manager of the DFN project LetsQoS. In May 2004, he joined the Research and Development Department, Siemens, Corporate Technology, where he became Senior Research Scientist. In October 2006, the department was transferred to Siemens Networks, which then has become part of the joint–venture Nokia Siemens Networks (NSN) in April 2007. At NSN, he became project manager and the responsible person for control plane topics (Layer2 – Layer4).

Andreas Kirstädter received the degree in electrical and industrial engineering from Munich University of Technology (TUM), Munich, Germany, in 1990 and 1992, respectively. From 1991 to 1997, he was with the Institute for Communication Networks, TUM, where he received the Ph.D. degree in the area of high-speed networking, WDM networks, simulation, analytical modeling, and high-speed hardware design.

In 1997, he joined Siemens Corporate Technology, Munich, where he was leading the high-speed and optical networks research. Currently, he is responsible for the Network Control-Plane and Transport Department, Nokia Siemens Networks GmbH and Co. KG, Munich. His research interests include architectures and protocols for the future Internet, carrier-grade Ethernet, photonic networks, control-plane architectures, hardware implementation of communication protocols, network resilience, mobility, QoS, and multimedia concepts for the Internet. He has been lecturing on "Simulation of Communication Networks" at TUM since 2000.

Biswanath Mukherjee (S'82–M'87–F'07) received the B.Tech. (Hons) degree from the Indian Institute of Technology, Kharagpur, in 1980. He received the Ph.D. degree from the University of Washington, Seattle, in June 1987, where he also held a GTE Teaching Fellowship and a General Electric Foundation Fellowship.

In July 1987, he joined the University of California, Davis (UC Davis), where he has been Professor of Computer Science since July 1995 (and currently holds the Child Family Endowed Chair Professorship), and served as Chairman of the Department of Computer Science during September 1997 to June 2000. To date, he has graduated nearly 25 Ph.D. degree students, with almost the same number of M.S. degree students. Currently, he supervises the research of nearly 20 scholars, mainly Ph.D. degree students, including visiting research scientists in his laboratory. He is the author of the textbook *Optical WDM Networks* (New York: Springer, January 2006). Earlier, he authored the textbook *Optical Communication Networks* (New York: McGraw-Hill, 1997), a book which received the Association of American Publishers, Inc.'s 1997 Honorable Mention in Computer Science. His research interests include lightwave networks, network security, and wireless networks.

Dr. Mukherjee is the winner of the 2004 Distinguished Graduate Mentoring Award at UC Davis. Two Ph.D. degree dissertations (by Dr. L. Sahasrabuddhe and Dr. K. Zhu), which he supervised, were winners of the 2000 and 2004 UC Davis College of Engineering Distinguished Dissertation Awards. He is a co-winner of paper awards presented at the 1991 and the 1994 National Computer Security Conferences. He serves or has served on the editorial boards of the IEEE/ACM TRANSACTIONS ON NETWORKING, IEEE *Network*, *ACM/Baltzer Wireless Information Networks (WINET)*, the *Journal of High-Speed Networks*, *Photonic Network Communications*, *Optical Network Magazine*, and *Optical Switching and Networking*. He served as Editor-at-Large for optical networking and communications for the IEEE Communications Society; as the Technical Program Chair of the IEEE INFOCOM '96 conference; and as Chairman of the IEEE Communication Society's Optical Networking Technical Committee (ONTC) during 2003–2005. He is a Member of the Board of Directors of IPLocks, Inc., a Silicon Valley startup company. He has consulted for and served on the Technical Advisory Board (TAB) of a number of startup companies in optical networking. His current TAB appointments include: Teknovus, Intelligent Fiber Optic Systems, and LookAhead Decisions Inc. (LDI).