A New Approach for Automatic Grooming of SONET Circuits To Optical Express Links

I. Widjaja, I. Saniee, L. Qian, A. Elwalid, J. Ellson and L. Cheng
Bell Laboratories, Lucent Technologies
Murray Hill, NJ 07974

Abstract—**We consider meshed optical transport networks having multiple levels of hierarchy whereby cross-connects in different levels perform multiplexing and demultiplexing functions at different granularities. We present a traffic grooming approach that can be implemented in a centralized or distributed fashion, based on the novel design that takes advantage of algorithm efficiency and a simple threshold mechanism to decide when grooming is economical. Our approach allows nodes to perform grooming and degrooming automatically, which is desired in the next-generation optical network where circuits are setup and torn-down dynamically through signaling. We present several experiments using different network topologies. The results indicate that the flow pattern influences the port requirement behavior, and that the flow thickness influences the optimal value of the threshold.**

Keywords—**Optical network, grooming, cost optimization, centralized and distributed algorithms.**

I. Introduction

Next-generation optical networks are expected to be capable of multiplexing many wavelengths in a single fiber where each wavelength typically carries high-bandwidth traffic (e.g., OC-192, OC-768, etc.). However, since typical services to end customers still consist of many relatively slow-speed circuits (e.g., DS-3, OC-3, etc.), cross-connects that can switch at the smallest granularity of the end-to-end services are still required. A naive solution is to require *all* cross-connects in the optical network be capable of switching at the smallest granularity. Unfortunately, the cost of the optical network could be prohibitively expensive if all slow-speed circuits need to be demultiplexed and multiplexed at each cross-connect. For instance, if a cross-connect having 1024 OC-192 ports needs to handle down to STS-1 circuits, then the cross-connect has to be capable of switching as many as 196,608 circuits. Clearly, such a large cross-connect is wasteful if there are many groups of STS-1 circuits transiting the cross-connect from the same input ports to the same output ports - a situation that typically occurs at intermediate nodes where there is already a substantial amount of traffic aggregation.

A more cost-effective optical network can be viewed as having multiple levels of hierarchy whereby cross-connects in different levels perform multiplexing and demultiplexing functions at different granularities. A cross-connect that performs local drops and adds usually needs to handle circuits at the lowest granularity (e.g., at the STS-1 level). Sufficiently aggregated traffic between two distant cross-connects may be passed to a different level of hierarchy with cross-connects that switch at the higher granularity (e.g., at the wavelength level). At the other extreme, highly aggregated traffic connecting between two cities may be switched at the fiber level. The process of multiplexing circuits and transporting the aggregate as a single unit to a common node where the aggregate is eventually demultiplexed is called *traffic grooming*. Fig. 1 shows a particular example of an optical network with two levels of hierarchy consisting of digital cross-connects (DXCs) capable of switching at STS-1 granularity at one level, and photonic cross-connects (PXCs) capable of switching at STS-192 granularity at another level (in practice, adjacent DXCs/PXCs are connected through DWDM systems that are not shown in the figure). In this example, the aggregate traffic between nodes A and D is high enough that it may be worthwhile to transport the aggregate directly via an "optical express link" (OEL) between nodes A and D facilitated through a set of PXCs (in the example, through nodes I and J). Other circuits that pass through nodes A and D but terminated at different nodes may also be groomed through the same OEL. It should be noted that our concept of optical express link can be generalized to an arbitrary link operating at any mode (TDM, lambda, fiber, etc.).

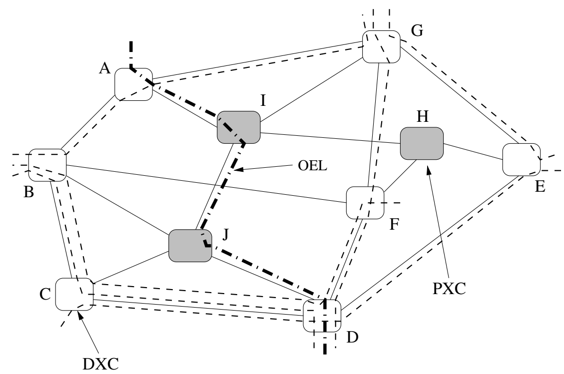

Fig. 1. Optical network with two levels of hierarchy.

In this paper, we present a new approach for traffic grooming and degrooming that can be implemented in a centralized or distributed fashion. Indeed, a distributed grooming algorithm is gaining significant importance in light of recent advances in automatic-switched optical networks such as GMPLS [1]. Our novel design takes advantage of algorithm efficiency and a simple threshold mechanism to decide when grooming or degrooming is economical.

The general subject of traffic grooming has been explored by several researchers in the past. Much of previous work (e.g., [2], [3], [4]) focuses only on ring networks. MacGregor and Grover [5] consider transport networks of arbitrary topology. However, their grooming approach only allows circuits that travel through the same sequence of spans. On the other hand, our approach, which does not impose this restriction, appears to be increasingly relevant in light of recent traffic-engineered net-

0-7803-7802-4/03/$17.00 © 2003 IEEE

works that often allow multiple paths even between the same source-destination pairs. Note also that [5] does not consider a distributed implementation.

The rest of this paper is organized as follows. Sec. II describes the traffic grooming problem formulation, and the decision used for trigerring the grooming or degrooming process. Sec. III details both the centralized and decentralized versions of our grooming approach. We illustrate the results via several experiments in Sec. IV using ring and meshed topologies. Finally, we conclude the paper in Sec. V.

II. Optical Express Links

A. Problem Formulation

Let the granularity of the end-to-end circuit be at the STS-n level. In general, only existing end-to-end circuits that are "long-lived" (e.g., holding time of hours or days) would be identified as candidates for grooming. The network is considered to be of arbitrary topology and consists of two levels of hierarchy: one level with DXCs capable of switching STS-n circuits and another level with PXCs capable of switching STS-N circuits (Fig. 1 shows an example where $n = 1$ and $N = 192$). The overall objective is to move portions (or segments) of end-to-end STS-n circuits into OELs established through PXCs so that the total network cost is minimized. The circuits traversing two common DXCs that are moved to an OEL do not have to follow the same route or terminate at the same source or destination. Because some traffic disruption may be encountered, traffic grooming may be preferably done during times of light usage (e.g., at night). Moreover, the establishment of OELs should occur at intervals much larger than the time it takes to reroute STS-n circuits to new OELs to minimize disruption. When end-to-end STS-n circuits are torn down, it is possible that the utilization of some existing OELs may become very low. In this case, it may be more economical to perform degrooming to revert to the original non-OEL routes (or to new non-OEL routes which are deemed to be better). Finally, it should be noted that our traffic grooming problem can be generalized from two levels to m levels.

B. Grooming Decision

We propose to use a simple threshold mechanism to determine when it is suitable to groom or degroom circuits. In particular, a group of k STS-n circuits traversing two common DXCs are eligible to be groomed to an OEL if the utilization, $\rho = kn/N$, is greater than or equal to the threshold θ. For example, suppose that STS-3s are to be groomed into STS-192s (n=3 and N=192) with $\theta = 0.75$. Then there should be at least 48 STS-3s to justify a new OEL. If $k > N$, then multiple fully-utilized OELs may be built and the remaining circuits are eligible if the remaining utilization is greater than θ.

The value of the threshold may be supplied in different ways. It may be provisioned by the network operator if circuits are relatively static. If circuits are relatively dynamic, a more sophisticated way is to automate the decision process by having an offline system perform a modeling analysis and calculate the network cost function parameterized by the threshold value. The optimal value of θ corresponding to the optimal network cost can then be flooded throughout the network (e.g., via a routing protocol such as OSPF). This approach allows each DXC to automatically initiate the degrooming process. Of course, care must be taken so that some coordination among different DXCs are taken into account. Our distributed algorithm described later provides such an implementation.

In the same way that new OELs need to be *set up* due to addition of new circuits in a network, old OELs may need to be *taken down* due to changes or reduction of circuits in a network. To this end, we describe two procedures: one for *adding* new OELs and one *deleting* old OELs. An OEL will be torn-down and the associated circuits moved to non-OEL routes if the utilization ρ is less than or equal to $\hat{\theta}$. The threshold $\hat{\theta}$ should be smaller than θ so that oscillation will not happen between the add and delete procedures. Moreover, sufficient hysteresis is generally desirable to ensure that small churns would not precipitate erratic grooming and degrooming.

III. Algorithms

The optimal mathematical solution to the grooming problem involves simultaneous establishments of multiple OELs. The solution technique typically requires integer programming that is: 1) centralized, 2) often hard to scale for large problems, and 3) computationally intensive. Instead, we propose an approximation scheme which is efficient and naturally lends itself to a distributed implementation.

A. Centralized algorithm

The centralized version utilizes the available network state: the information about the network topology, existing capacities, circuits, routes, and existing and available ports. The state information can be kept in a centralized database such as a Network Management Station (NMS). Periodically, the optimization algorithm can be performed to setup/tear-down OELs as necessary.

Before describing the algorithm, let us first introduce some notations. Let flow f represent a group of end-to-end bidirectional STS-n circuits having the same termination nodes, and F be the set of all flows. Assume that each flow is bidirectional. The bandwidth of flow f is given as $v(f)$, measured in units of STS-n's. For brevity, we say that flow f contains cross-connect x if f passes through cross-connect x. Let E be the set of all links in the network that connect two DXCs directly, and $len(l)$ be the metric (e.g., length) on link l in E. Unlike a direct link that is permanent, an OEL is a link connecting two DXCs via one or more PXCs, and can be setup or torndown dynamically. Let E_{OEL} denote the current set of OELs. Define $B(x, y)$ as the current total bandwidth of the flows that pass through DXCs x and y, and $B_l(x, y)$ as the current total bandwidth-length metric between DXC x and DXC y. Both $B(x, y)$ and $B_l(x, y)$ are set 0 before the algorithm is run. Define $S_{OEL}(x, y)$ as the spare capacity of OELs that either exist or can be built between DXC x and DXC y. Let $P(x, y; f)$ represent the sequence of links in E followed by flow f between DXC x and DXC y, and $len(l)$ be the length (or general metric) on link l. Then the centralized algorithm that adds OELs is described as follows.

Scheme ADD_OEL (centralized):
repeat until no more OEL can be generated
for each DXC pair (x, y) not in E
for each $f \in F$ that contains (x, y)
if x and y have available ports for OELs
if $B(x, y) + v(f) <= S_{OEL}(x, y)$
$B(x, y) += v(f)$
$B_l(x, y) += \sum_{l \in P(x,y;f)} v(f) * len(l)$
else
break and pick the next pair
if $B(x, y) \geq \theta$
set $OEL(x, y) = B_l(x, y) - B(x, y)$
let $(x*, y*)$ be the pair with maximal $OEL(x, y)$
update_add$(x*, y*)$

Note that ties among pairs with maximal OELs can be broken randomly. The procedure that updates flows from an old segment to a new OEL is given as follows.

update_add$(x*, y*)$
for each flow $f \in F$ that contains $(x*, y*)$
re-route f between $(x*, y*)$ via OEL so that
$f = (n_1, \cdots, n_i(x*), n_{i+1}, \cdots, n_j(y*), \cdots, n_z)$
becomes
$f = (n_1, \cdots, n_i(= x*), n_j(= y*), \cdots, n_z)$

Note that the OEL between $x*$ and $y*$ may physically traverse through a number of PXCs, but is logically treated as a link at the STS-n level. Note also that the running time of the algorithm for each OEL addition is $O(|F| \times |V|^2)$, where $|F|$ is the number of flows and $|V|$ is the number of DXCs.

The centralized algorithm that deletes OELs is described as follows.

Scheme DELETE_OEL (centralized):
repeat until no more OEL can be deleted
for each DXC pair (x, y) in E_{OEL}
for each $f \in F$ that contains (x, y)
if $B(x, y) < \hat{\theta}$
add oldFlow(x,y) to OEL removal list
let $(x*, y*)$ be pair with minimal oldFlow(x, y)
update_delete$(x*, y*)$

In the above description, oldFlow(x, y) tracks the information of flows that contain (x, y). The procedure that updates flows from an OEL back to the new non-OEL segment is given as follows.

update_delete$(x*, y*)$
for each flow f through OEL containing $(x*, y*)$
in order of non-increasing bandwidth
find the constraint shortest path for f between
$x*$ and $y*$ which avoids the OEL
if such a path is available
re-route f

If flows being moved from an OEL results in the OEL being empty, then the OEL will be torn-down. The overall scheme involves first running the DELETE_OEL algorithm which deletes as many OELs as necessary until no more removals are possible, then executing the ADD_OEL algorithm until no further OEL additions are possible.

B. Distributed algorithm

The main idea is to emulate the behavior of the centralized algorithm and not rely on a centralized database. To this end, each DXC x computes the best OEL that it can find to node $y*$. It then broadcasts the computed $OEL(x, y*)$ to the rest of the network. Other DXCs v with $OEL(v, w)$ less than $OEL(x, y*)$ drop out. Ultimately the DXC with the highest OEL value wins. Ties are broken by the DXC address (e.g., DXC with the highest address wins the tie-breaker).

The most basic assumption is that each DXC maintains the route of each active incoming or outgoing STS-n circuit that goes through it. Let $\omega(x, f)$ be the set of DXCs including x that are contained by flow f. Let $\Omega(x) = \bigcup_{f \in F} \omega(x, f)$. For brevity, we only present the technique to add OELs in this paper. The distributed algorithm that adds OELs is described as follows.

Scheme ADD_OEL (decentralized)
each DXC x independently does as follows
for each $y(\neq x)$ in $\Omega(x)$
for each flow f that contains (x, y)
$B(x, y) += v(f)$
$B_l(x, y) += \sum_{l \in P(x,y;f)} v(f) * len(l)$
if $B(x, y) \geq \theta$ and x can set up an OEL to y
set $OEL(x, y) = B_l(x, y) - B(x, y)$
let $y*$ be the DXC with maximal $OEL(x, y*)$
broadcast $OEL(x, y*)$ to all DXCs in network
if x receives and $OEL(v, w) > OEL(x, y*)$
stops broadcasting and refrains from setting
up an OEL
updates all flows through it based on any
$OEL(v, w)$ received
else
broadcasts the optimal $OEL(x, y*)$
set up an OEL to $y*$
redirect existing flows to the new OEL
broadcasts the new $OEL(x, y*)$ and $y*$

IV. Experiments

In this section, we examine how the optimal value of θ (correspondingly $\hat{\theta}$) can be obtained for a particular environment. As mentioned previously, the total network cost which is parameterized by θ is the objective function to be minimized. Once the minimum network cost is determined, the corresponding value of θ can be loaded to each DXC so that the grooming algorithm can operate efficiently in a distributed manner. Of course, the value of θ may need to be updated periodically to respond to changes in demand patterns.

In general, the network cost depends on many factors such as the cost of an OEL, the DXC port cost, the PXC port cost, the cost of fiber/wavelength, the numbers of hops between two DXCs (x, y) where traffic is being groomed along the original route and the OEL route, and other quantities that have appreciable impact on the network cost. Although our approach works for any type of cost function, we choose a simple yet realistic cost function that captures the modularity of network cost. In particular, the cost function is given by

$$Total\ cost = N_{STS1} \times C_{STS1} + N_{STS192} \times C_{STS192},$$

where N_{STS1} is the number of DXC channelized STS-192 ports and C_{STS1} is the cost of each DXC channelized STS-192 port. Similarly, N_{STS192} is the number of PXC unchannelized STS-192c ports and C_{STS192} is the cost of each PXC port. Note that switching and other costs are assumed to be absorbed by the port costs.

Fig. 2 shows an example of the network cost assuming that f_1 STS-1 circuits are to be routed between DXC 1 and DXC

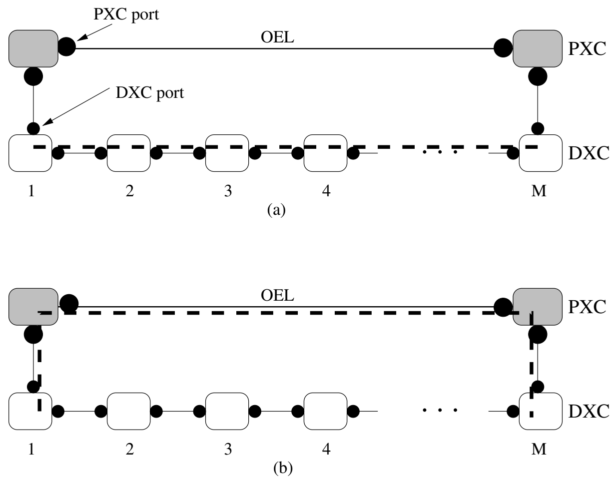

Fig. 2. Total costs for (a) ungroomed and (b) groomed circuits.

M. Without grooming (see Fig. 2a), the total number of DXC ports from the above model is $2(M-1)\lceil f_1/192 \rceil$, and the total cost is $2(M-1)\lceil f_1/192 \rceil \times C_{STS1}$. With full grooming (see Fig. 2b), the total number of DXC ports is $2\lceil f_1/192 \rceil$, while the total number of PXC ports is $4\lceil f_1/192 \rceil$. Thus the total cost is $2\lceil f_1/192 \rceil \times C_{STS1} + 4\lceil f_1/192 \rceil \times C_{STS192}$.

A. *Ring Network*

We first consider a ring network which is typically deployed in a metro environment. The network consists of 14 DXCs, and the flows from any DXC are uniformly distributed to all other DXCs that are greater than two hops away. We assume that each flow has bandwidth of 10 STS-1s. We also assume that an OEL can always be built from any DXC to any other DXC, if needed.

Fig. 3 plots the total number of ports required for DXCs and PXCs as a function of θ. As can be seen from the figure, the number of ports required for PXCs decreases as θ increases. This is to be expected since a higher value of θ translates to less opportunity to build OELs. Notice that the opposite holds true for DXCs in this scenario.

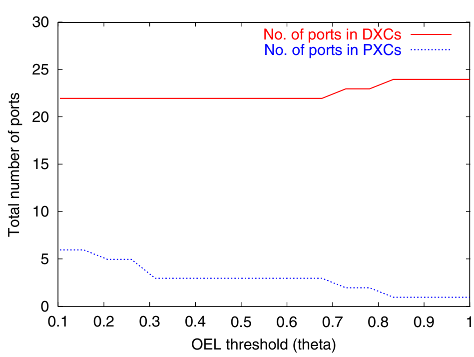

Fig. 3. Port requirement as a function of θ for a ring network.

Fig. 4 plots the total network cost as a function of θ for different values of R, defined as $R = C_{STS1}/C_{STS192}$ (note that R is much larger than 1 in practice.). For fixed values of C_{STS1} and C_{STS192}, the total costs can differ significantly for different values of R. For comparison purpose, we adjust C_{STS1} and C_{STS192} appropriately for a given value of R so that the total costs for different values of R are approximately within the same range, and therefore we refer the costs to be *relative*. Notice that when $R = 2, 5, 10$, the optimal threshold value approximately lies from 0.3 to 0.7. This result allows us to set θ to 0.7 and $\hat{\theta}$ to 0.3. Notice also that when $R = 1$, the plot indicates that it is always no worse not to build an OEL independent of the utilization ρ of a particular segment.

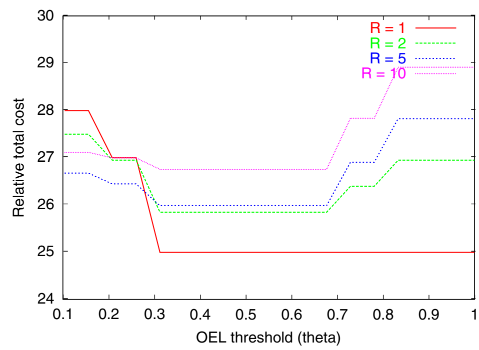

Fig. 4. Total cost as a function of θ for a ring network.

It is worthwhile to point out that available analytic lower bounds for ring networks (e.g., [2]) are not applicable under our more realistic assumptions, especially due to the modularity of channelized port cost that we allow in our cost function. Under our cost model, there is no known simple analytical quantification of grooming in ring networks even under the assumption that the demand is uniform, since the model is at least as hard as known combinatorially hard problems (e.g., bin packing [6]).

B. *Meshed Network*

We now turn our attention to another network topology which is representative of a core U.S. networks, as shown in Fig. 5. The network contains 31 DXCs (the PXCs are not shown in the figure).

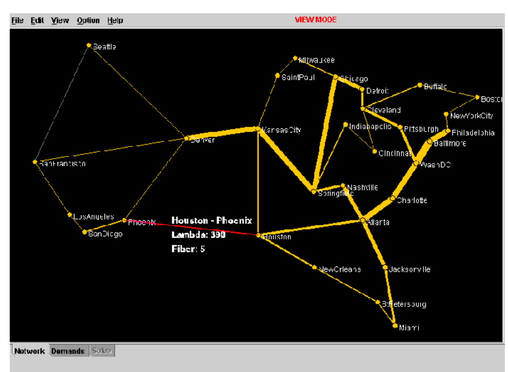

Fig. 5. Generic U.S. network.

To create the traffic demands, we first generate a flow from one DXC to any other DXC according to a uniform distribution from 0 to 6 STS-1s, giving a mean of 3 STS-1s. We repeat this flow generation process from other DXCs taking into account that if a flow has been generated from DXC i to DXC j previously, then no action will be taken from DXC j to DXC i since a flow is already assumed to be bi-directional.

Figs. 6 shows the port requirement as θ is varied. Notice that the port requirement in Fig. 6 is different from that in Fig. 3 in the sense that the number of ports in DXCs actually increases as θ becomes very small. This is because the present example has a significant number of end-to-end circuits between adjacent DXCs, resulting in ungroomable flows. Thus a certain number of ports in DXCs are required no matter how low θ is. On the other hand, a low value of θ which results in a high number of OELs requires additional ports in DXCs to connect to the PXCs. We also observed this phenomenon in ring networks.

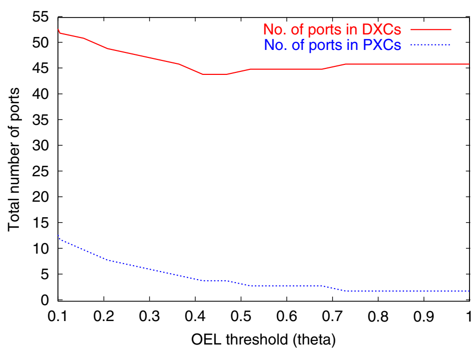

Fig. 6. Port requirement as a function of θ for a meshed network.

Fig. 7 displays the corresponding cost function. For $R = 2, 5$, and 10, notice that the optimal threshold value now indicates that θ should be set to about 0.5 and $\hat{\theta}$ to about 0.4. Observe that this example allows for a smaller hysteresis than the previous one.

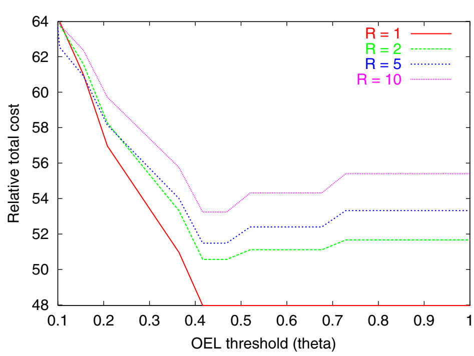

Fig. 7. Total cost as a function of θ for a meshed network.

Figs. 8 and 9 show the scenario where the flows are significantly "thicker" than before. In particular, the flow between each pair of DXCs is determined according to a uniform distribution from 10 to 15 STS-1s. As can be verified from Fig. 8, this scenario requires more ports to route all the demands. A more interesting observation is revealed in Fig. 9 where the optimal threshold value is now closer to full utilization.

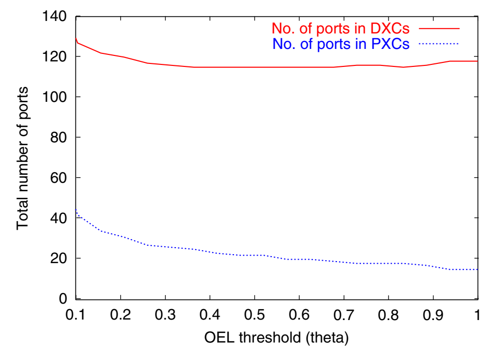

Fig. 8. Port requirement as a function of θ with higher demands.

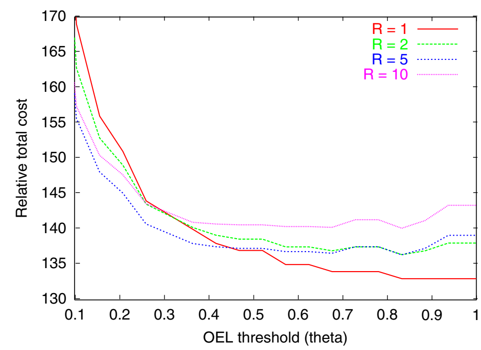

Fig. 9. Total cost as a function of θ with higher demands.

V. Conclusion

We have presented an approach for traffic grooming and degrooming that can be implemented in a centralized or distributed fashion. The centralized algorithm may be more suitable for optical networks that provision circuits via a central station such as an NMS, while the distributed algorithm may be more suitable for automated switched optical networks relying on signaling to provision circuits. The grooming and degrooming decisions are synthesized by a simple threshold mechanism, so that decision to setup or tear down an OEL can be made rapidly. Our experimental results indicate that the flow pattern influences the port requirement behavior, and that the flow thickness influences the optimal values of the threshold. Our present work can be extended in several directions. In particular, we are currently exploring an extension that incorporates protection circuits into the grooming algorithms.

References

[1] E. Mannie et al., "Generalized Multi-Protocol Label Switching (GMPLS) Architecture," IETF Draft, Work in Progress, Jun. 2001.

[2] A. G. Chiu and E. .H. Modiano, "Traffic grooming algorithms for reducing electronic multiplexing costs in WDM ring networks," *IEEE Journal of Lightwave Technology*, vol. 18, no. 2, pp. 2–12, Jan. 2000.

[3] O. Gerstel, R. Rawaswami, and G. Sasaki, "Cost-effective traffic grooming in WDM rings," *IEEE Trans. on Networking*, vol. 8, no. 5, pp. 618–630, Oct. 2000.

[4] X. Zhang and C.. Qiao, "An effective and comprehensive approach for traffic grooming and wavelength assignment in SONET/WDM rings," *IEEE Trans. on Networking*, vol. 8, no. 5, pp. 608–617, Oct. 2000.

[5] M. H. MacGregor and W. D. Grover, "Distributed partial-express routing of broadband transport network demands," *IEEE Trans. on Networking*, vol. 5, no. 6, pp. 981–988, Dec. 1997.

[6] M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, Freeman, New York, 1979.