

# Influence of Multilayer Traffic Engineering Timing Parameters on Network Performance

B. Puype, D. Colle, M. Pickavet, P. Demeester  
Dept. of Information Technology, University of Ghent – IBBT – IMEC  
Gaston Crommenlaan 8 bus 201, B-9050 Gent, Belgium  
E-mail: {bpuype; dcolle; mpick; demeester}@intec.ugent.be

**Abstract-** Recent advances in optical networking technology have moved the state-of-the-art from manually installed fiber connections to fully automatic switched lightpaths. Multilayer Traffic Engineering (MTE) in an IP-over-Optical network allows to leverage rapid lightpath setup/teardown as a cross-layer traffic engineering technique. It enables on-the-fly reconfiguration of the IP layer logical topology and up/downgrade of the capacity of IP links. Together with classical IP layer routing techniques, MTE intelligently solves problems such as IP layer congestion and packet loss and it may optimize optical layer capacity usage and total network throughput. In this, the rate at which MTE can make adjustments to the network is limited by technology and stability concerns. We present some example MTE techniques and discuss how the timing parameters of these mechanisms impact perceived network performance.

## I. INTRODUCTION

Multilayer traffic engineering (MTE) is a form of traffic engineering that utilizes optical layer switching technology as an asset to increase upper layer network performance in terms of throughput, traffic loss, etc. In addition to the routing of traffic flows, which is the traditional goal of traffic engineering, the more complex aspect of MTE is the design of a logical topology for upper network layers using the underlying optical layer as bandwidth provider. For an IP-over-Optical network for example, IP routers are connected using optical connections (lightpaths), thus forming an IP layer logical topology that can be completely different from the underlying physical topology (optical fibres and switches).

The problem is such that the routing of IP traffic interacts with the logical topology design itself: routing requires a topology, and adjustments to that topology are usually triggered by certain congestions, in term caused by routed traffic. Therefore traffic demands that vary over time may be coped with through either routing traffic over other paths, or reconfiguring the logical topology (adding capacity, for example).

Nevertheless, when looking at topology design, two extreme cases can be identified. On the one hand, there is point-to-point grooming, where one will rely mostly on creating (long) multi-hop paths for traffic flows, relying on a sparser logical topology, sometimes closely resembling the physical one. This method will emphasize the routing aspect, and put the burden on the routing capabilities of the IP nodes. On the other hand, there is end-to-end grooming, preferring to establish (long) lower layer connections (lightpaths) instead, directly connecting the IP routers. In this case the routing

aspect is less important (much of the traffic will have one-hop paths). These types of topologies are mostly supported by the optical layer (and large availability of IP-optical interfaces). As a consequence, bandwidth efficiency of optical connections may suffer, since it becomes harder to arrange traffic flows such that they ‘fill up’ lightpaths.

In [1], we explained that one way to classify MTE strategies is to separate them into reactive and proactive algorithms. The dynamic behaviour of reactive strategies is determined by the fact that they trigger network optimization on certain events. Typically this can be detection of congestion, a new incoming traffic flow, a traffic measurement, etc. Triggers may be acted on immediately, or they may be delayed until a sufficient amount of change is seen; for example, a MTE action can be initiated only after the measured traffic volume change of a traffic flow exceeds a pre-determined threshold.

For the proactive approach on the other hand, MTE optimization is performed continuously. For a practical implementation, this may mean that the MTE optimization process is performed (‘triggered’) periodically. Since in this case there is no explicit initiating cause (except e.g. a timer elapsing), algorithms of proactive MTE strategies will be such that the entire network is considered. A single optimization cycle may affect routing or topology changes for multiple traffic flows. In contrast, for the reactive case, the problem is identified on beforehand, so it suffices to limit the search-space of the algorithm to the problem area (e.g. to the flow whose bandwidth fluctuation triggered the optimization).

When conceiving a MTE strategy, a compromise needs to be made between optimality and the rate at which MTE actions are performed. Although network throughput, loss, etc. can in theory be held optimal through very frequent rerouting, bandwidth adjustment and logical topology reconfiguration, practical issues will limit the rate of such actions. While lightpath setup and teardown can be performed on a sub-second timescale, the data plane (typically TCP/IP) will suffer from frequent changes to routing tables and capacity rearrangement. Also, such routing changes trigger high volumes of routing messages which are flooded through the network, in order to inform other routers of connectivity and keep this information in all network nodes synchronized.

Therefore, certain limitations exist on the rate at which such MTE action may be performed. For the reactive case, this may be controlled by delaying the triggers (e.g. by introducing a triggering threshold, as mentioned before). For the proactive case, the frequency at which the MTE cycle is

initiated is one of its natural parameters. However, one optimization cycle may trigger none, some or even many separate MTE actions (lightpath setup/teardown, rerouting...). In both reactive and proactive cases, one can expect this rate to stabilize during normal operation, or at least to be directly related to the rate of change seen in the performance metrics that are being optimized by the algorithm (be it network throughput, traffic loss, QoS...).

## II. Multilayer Traffic Engineering strategy

In this paper, we will examine the proactive case for a specific MTE algorithm which was first presented in [1]. This MTE strategy acts upon traffic flows on a conceptual level. These flows represent and model non-elastic traffic through a small number of parameters, the most important one in our case being flow bandwidth. The traffic is therefore not modelled up to the detail of single packets, and so very small timescales are not considered. The discussion of the MTE algorithm will be based on simulations using mathematically generated flow bandwidths. However, one may conceive these flow bandwidths to be abstracted from measurements on an operational network so that in fact the presented MTE strategy could be used to configure an actual network instead of a simulated one.

The network scenario envisioned is an IP/MPLS over Optical two-layer network stack, where the flows from the model correspond to IP/MPLS LSPs, and the logical topology is formed using optical (e.g. DWDM) lightpaths. The objective of the MTE strategy during each optimization cycle is to take into account offered flow traffic (fluctuating in time) and to derive suitable LSP routes, IP/MPLS logical topology and required capacity of each of the IP/MPLS logical links.

### A. Cost function

Most of the algorithmic complexity of the MTE strategy was contained in a cost function to be used as a shortest path routing cost in the IP/MPLS layer (i.e. to establish explicit LSP routes). The cost function by itself will route and groom IP traffic such that lightpaths are utilized optimally. While some limited optical metrics[2] are taken into account in constructing the cost function, optical layer topology, switching functionality and wavelength occupation is largely ignored by the MTE algorithm. The optical layer is seen as a cloud which provides on-demand connectivity between nodes. The algorithm therefore acts mostly in the IP layer and based on information from this layer, except for some metrics exported from the optical layer. This approach fits very well with the overlay model of an IP over Optical network, which does not require the operators of each layer to reveal their topology or otherwise sensitive information. Also, interlayer information exchange (and dependence upon it) is limited in this way.

While the optical metric is a simple value translated into a multiplicative factor of the shortest path cost function, the IP layer part of the cost function depends on IP/MPLS link load. The cost function in its entirety then serves to perform traffic

grooming but also topology configuration. The latter may seem surprising at first, since of course the cost function is defined only for IP/MPLS links which are at that time available in the logical topology. In order to let the shortest path routing algorithm result in paths over not yet established IP/MPLS links (and therefore have the ability to request link setup), we will route traffic flows over a virtual full mesh of the IP/MPLS network. This full mesh models the optical layer flexibility in offering any connectivity desired. The logical topology configuration then comes down to the reduction of this full mesh into a more sparsely meshed logical topology (which is the one to be kept set up in the actual network). Nevertheless, since the cost function depends on link load, we will extract this load from the actual network and use it on the full mesh abstraction during MTE routing (non-established IP link are given a load of 0).

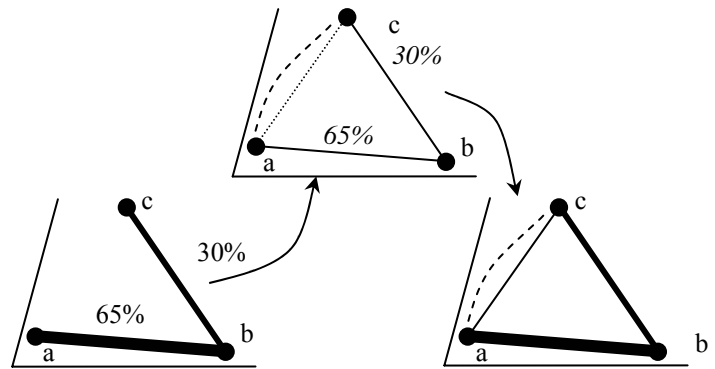


Figure 1. Virtual full mesh routing example

Figure 1 may clarify some of these concepts. Here we try to find a path for a traffic flow between nodes a and c. The actual logical topology established at first does not have a direct hop path between nodes a and c, so the only path initially available runs via node b. If we assume the shown link load on link a-b (65%) to be too high to allow additional flows over it, we may decide in the virtual full mesh routing step to instead route over the virtual link a-c (which is available in this virtual topology), although this will mean requesting a new lightpath from the optical layer, which is promptly done afterwards in order to effect this path in the actual logical topology.

This small example already hints at one more issue to keep in mind. Since the MTE strategy will route not only newly offered traffic flows, but also reroute (or at least, examine) flows which already have a path assigned to them, part of the load on the IP links may be caused by the flow in question. In this case, part (or even all) of the 65% load seen on link a-b may be due to traffic from the flow between a and c. This needs to be taken into account. Therefore, during the cost function calculation, the link load used is the load that a certain link *would* have if it were to carry said flow. For links in the current path of a certain flow, the load used is the measured one, for other links, it is the measured load + the flow load. Both flow bandwidth and current path are assumed to be known data (i.e. bandwidth of flows will be flooded through the network periodically).

The IP/MPLS part of the cost function then is characterised by three parameters. Firstly, there is a High Load Threshold – IP links with a load above HLT receive an exponentially rising cost. However, also lightly loaded links (with a load below Low Load Threshold, LLT) are penalised with a higher cost. The higher cost for low loads is defined against the cost for moderate loads by the Low/Moderate Ratio (LMR), indicating the ratio between cost for low loads (LC) and cost for moderate loads (MC);  $LMR = LC/MC$ . This cost penalty avoids establishing many and thus inefficiently used links, and thereby promotes grooming of traffic into IP links carrying a bundle of flows. The shape of the empirically designed cost  $C(L)$  was reached using the following equation:

$$C(L) = 2.LMR \left[ \exp\left(\frac{a(L-LLT)}{b}\right) + 1 \right]^{-1} + LMR \left[ \exp\left(\frac{-L}{ab}\right) + \exp\left(\frac{L-HLT}{ab}\right) \right] + 1 + 2.LMR \cdot \exp\left(\frac{L-HLT-0.15}{b}\right). \quad (1)$$

Load  $L$  is in  $[0, 1]$ . There are two shaping parameters ( $a, b$ ) = (4, 0.05) for all further results, which control curvature of the function. Important is the fact that the cost for overloaded IP links always dominates over the cost for lightly loaded links, because avoiding overload is more important than forcing grooming.

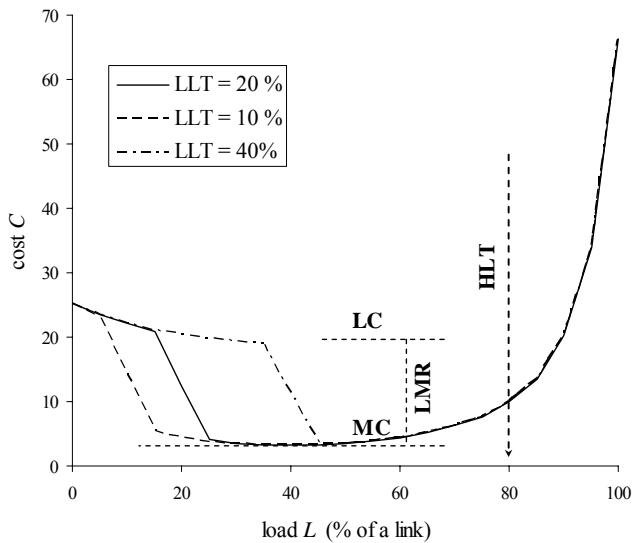


Figure 2. IP link load dependence of cost function

Figure 2 shows some example cost functions, where  $HLT = 80\%$  (anticipating packet loss for IP links loaded higher than this). A typical value of LLT would be 20%, but some variety of cost functions with different LLTs are shown. The LLT parameter will turn out to be important later on, because it influences the extent of the optimization done by the MTE strategy. Moderately loaded IP links have a utilization somewhere around 50%. Increasing LLT will reduce the ‘hole’ in the cost function, reducing the range of link loads which are deemed acceptable.

### B. Traffic flow bandwidth

The bandwidth of traffic flows during simulations discussed later on, consists of two components which vary over time. Each node pair in the simulated network is associated with two unidirectional traffic flows. The largest part of the traffic volume between node pairs is random uniformly distributed as detailed in [2]. The collection of generated traffic flows at a certain point in time is a traffic pattern (a matrix over all nodes). A random uniformly distributed traffic pattern can be characterized by a simple parameter, the maximum IP traffic of a single flow. For example, a pattern generated with a maximum traffic per flow of 10 Gbit/s, will result in a traffic matrix whose elements are between 0 and 10 Gbit/s (with an average of 5 Gbit/s). This maximum traffic per flow will be denoted  $B$  in the charts further on as an indication of traffic volume.

The second component consists of a smaller amount of traffic which is Poisson process generated, and acts as faster varying additional traffic on top of the bulk uniform traffic, which will allow to examine the MTE strategy’s performance for traffic variances between optimization cycles.

Timescales for MTE cycles and bulk/additional traffic variations will be related through a parameter  $T_{bulk}$ , giving the number of timesteps between updates to the bulk traffic pattern. The Poisson process generated traffic’s bandwidth for each flow in fact results from a Markov chain for tractability. The discrete time Markov chain uses the above mentioned timestep as time parameter. In other words, increasing  $T_{bulk}$  will create larger stretches of time between bulk volume updates, and therefore more and larger variations to the additional traffic component, which is furthermore characterized by its  $\lambda$  and  $\mu$  parameters (Markov chain) or its average bandwidth in Erlang.

Because updates to the bulk traffic incur large changes in traffic, the MTE optimization cycle will be triggered in unison with the updates. Therefore  $T_{bulk}$  will also describe the time  $T_{MTE}$  between logical topology and routing updates. Both parameters will henceforward be denoted  $T$ . Note that logical topology changes coincide with routing updates in this strategy, since the topology is dictated by the paths assigned to traffic flows.

### III. Logical topology updates

In this section, we will discuss the impact of the time  $T$  between logical topology updates (and thus MTE optimization cycles). Increasing  $T$  will allow the total traffic to deviate further from the measured volume (bulk + additional traffic) that was used during the previous optimization cycle. This may cause some of the groomed traffic to exceed the available IP/MPLS capacity, leading to loss of traffic. As said, traffic is considered on a flow level, and not on a packet level. However, in [3] we examined the influence of timing parameters (e.g. measuring interval) with very small timescales on traffic modelled on the packet level.

Simulations were performed on the physical topology shown in Fig. 3. The optical layer is made up by 14 nodes interconnected by 23 optical fibre links. Typical optical

connection bandwidths are 2.5, 10, 40 Gbit/s. However, bandwidths in the IP layer will be denoted in percentages of a single lightpath, generalizing the results regardless of exact optical line system bit rate.

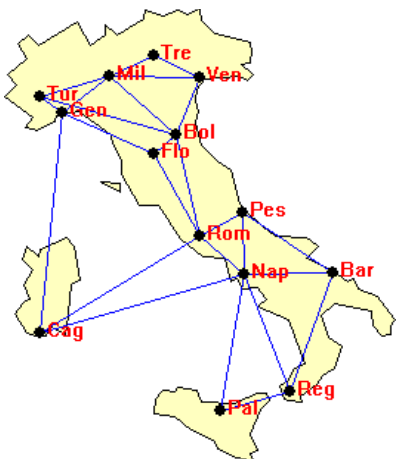


Figure 3. 14 node physical topology

For the bulk volume traffic pattern, we have assumed the maximum traffic per flow in a range from 5% to 55% (of a lightpath). The Markov chain process will generate additional traffic in the form of IP traffic requests of 5% bandwidth each, with a  $\lambda = 0.1$  and  $\mu = 0.05$ . Also, for this topology, at most one optical connection can be present between two IP routers, leaving the MTE strategy with logical topology design and routing, but not bandwidth adjustment on the logical IP links (see later on).

As simulation results, we are mostly interested in perceived loss of traffic in the IP layer. Although for real-life IP packet flows, one starts experiencing packet loss once the flow bandwidth exceeds around 80% of the line rate, we will assume that traffic loss occurs once the sum of all traffic flow bandwidths over an IP link exceeds that link's total capacity (i.e. 100%). This leaves loss calculation somewhat non-trivial. For example, for a single traffic flow, upstream loss will reduce load in downstream links, and thus possibly also their traffic loss. Equally, the loss rate for flows that share a same IP link will be correlated. This mutual interaction of loss rates leads to a complex system to be solved. As explained in [4], the loss rates for all links in a complex topology can be calculated using an iterative approach. The same method was used in this publication.

Fig. 4 shows traffic loss rates (averaged for all flows) in function of the maximum bandwidth per flow  $B$  of the bulk traffic pattern. The period  $T$  between routing and topology updates has its largest influence for traffic patterns with lower total volume. Partly, this is because the fast varying traffic component becomes a smaller component of the total traffic with increasing bulk volume. However, the varying component is taken into account during the logical topology design, so sufficient bandwidth is allocated for that additional traffic. Also, during simulations we experienced that loss rates remain relatively independent of the total bandwidth of the additional traffic component, except for very large

fluctuating traffic volumes (e.g. of the same order as the bulk traffic). In that case, loss rates shoot up to 10%.

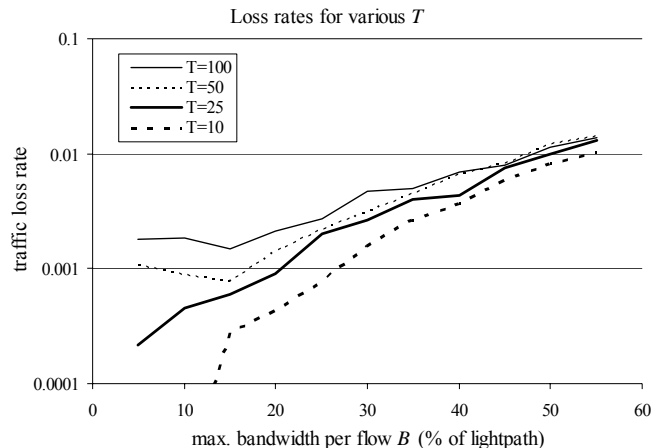


Figure 4. Influence of  $T$  on loss rates

The reduced influence of  $T$  for higher traffic volumes can be explained by the fact that very high traffic volumes will lead to a higher amount of meshing in the logical topology, in order to cope with the demand. We will start to see more end-to-end grooming (using long lightpaths). For low traffic volumes on the other hand, point-to-point grooming dominates, with many flows having multi-hop paths. For very low volumes, the logical topology may effectively be constructed as a spanning tree. Obviously, sparse topologies will be more specialized towards certain traffic patterns, while densely meshed topologies are less affected by traffic pattern changes, which explains the reduced reliance on topology reconfiguration for high volumes. Of course, in absolute figures, loss rates are the highest (1%) for high bandwidth volumes, since in those cases the network is operating at its limit.

We can further illustrate this behaviour by directly modifying the MTE algorithm so that it optimizes more aggressively towards either end-to-end or point-to-point grooming. This is done by changing the LLT parameter of the IP layer cost function (see Fig. 2). By increasing LLT (narrowing the optimal load range), the algorithm will try to even more avoid lightly loaded IP links, and route a higher amount of traffic flows on the remaining IP links. Decreasing LLT will loosen the soft constraints on the path selection, but therefore increase the required set of IP links from the virtual full mesh (thus promoting end-to-end grooming).

Fig. 5 and 6 show this influence for a range on LLT from 10% – 40%. For low and moderate LLT, loss rate in function of traffic volume is similar to the results from Fig. 4 (since LLT=20% is used for all other simulations). By increasing LLT, point-to-point grooming is forced. This causes an increase in traffic loss, especially for low traffic volumes. For high volumes, it cannot be forced because a sparse topology would not offer enough IP bandwidth (lightpath capacity) to cope with the offered traffic. We can now clearly see that loss is correlated with the extent of the point-to-point grooming in the logical topology.

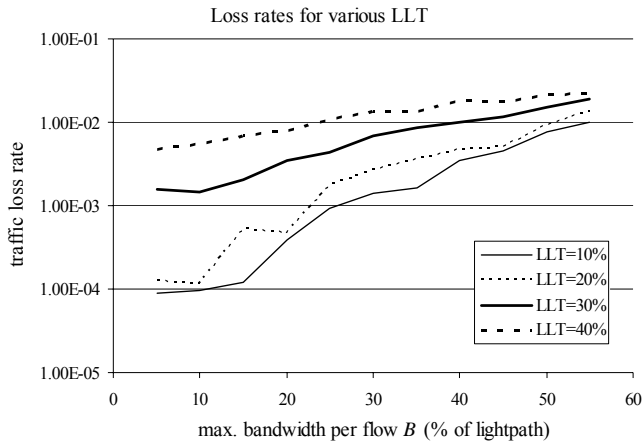


Figure 5. Influence of LLT on loss rate

One of the reasons to select the LLT parameter carefully is depicted on Fig. 6. Here the influence of LLT on the mesh size of the logical topology is shown for the same range of LLT. Increasing LLT (and point-to-point grooming) will generate logical topologies with a lower amount of IP links (for same traffic volumes). One of the reasons that point-to-point grooming increases loss, is that it will use lightpaths more efficiently. However, optimized topologies have a harder time to cope with traffic patterns that differ slightly, as said before.

Looking at the total reduction in logical IP links seen when adjusting LLT, one sees a fairly constant amount of IP links (20-30) removed for the entire traffic volume range. Of course, in relative number, the impact is much larger for the sparser meshes at low traffic volumes, again confirming the higher influence of LLT for lower bandwidth volumes.

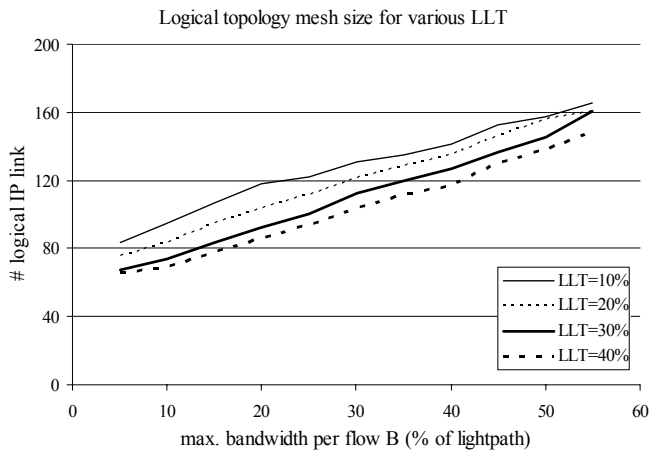


Figure 6. Influence of LLT on mesh size

To conclude this section, one may adjust the MTE parameters to reduce logical topology mesh complexity (and thereby optical layer capacity usage), but this creates more optimized and specialized topologies which possibly increase loss, depending on the traffic pattern. Some of these effects may be alleviated by increasing the optimization frequency (reducing  $T$ ). Though that in turn may lead to higher

instability or technical problems (e.g. given certain limitations on flooding rate).

#### IV. LINK UP/DOWNGRADE

As discussed in the previous section, very high traffic volumes cause the logical topology to tend towards high meshing. However, in large IP-over-Optical networks which are conceived for very high bandwidths, one may see an optical layer capable of setting up multiple parallel lightpaths between IP routers. With appropriate IP layer functionality, these can be used to construct very large bandwidth pipes between IP routers (or using the naming convention from this paper, logical links with a bandwidth of  $n \times 100\%$ , where 100% is a single lightpath). For large networks, this form of multilayer traffic engineering (IP link bandwidth up/downgrade) can be used in addition to logical topology reconfiguration (and routing), because it helps to reduce overall node degree and therefore IP router complexity and optical layer interface cost (though not router processing capacity).

In this section, we base ourselves on a different, much larger meshed pan-European topology with 28 nodes as seen on Fig. 7, which allows establishing up to 8 lightpaths at once between the same node pair.

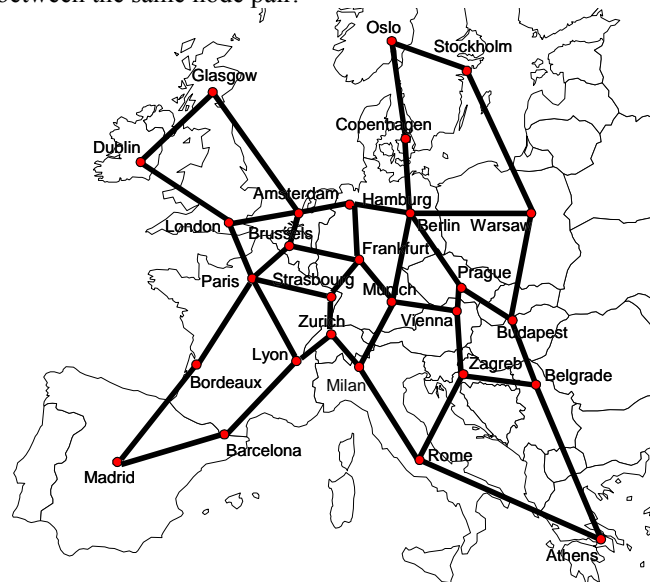


Figure 7. Pan-European 28 node physical topology

Being able to establish multiple lightpaths in parallel, and concatenating them into a single IP link, has as advantage that it is easier to stay in the point-to-point grooming regime, simply by upgrading IP link bandwidth as needed, without resorting to creating a very dense mesh. The previous section showed that, although traffic loss in the point-to-point regime has a larger variance, it is usually lower than loss experienced when shifting into the end-to-end grooming (dense mesh) regime. To adapt the MTE algorithm towards such a network scenario, we simply rescale the load axis of the cost function so that it spans the entire possible load range of our IP links (in this scenario, from 0 to  $8 \times 100\%$  of bandwidth).

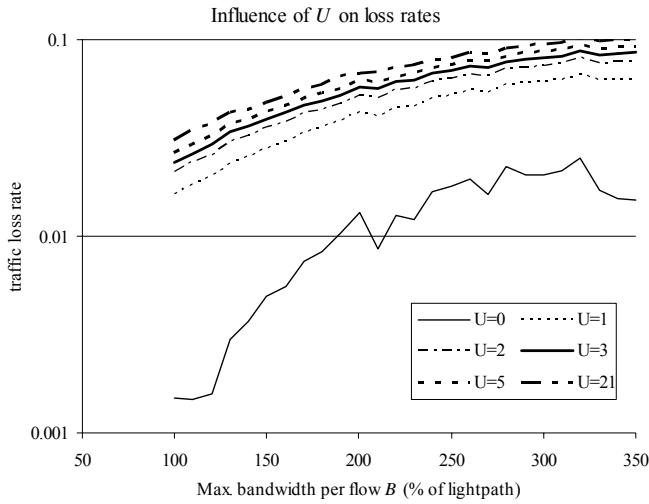


Figure 8. Influence of link up/downgrade process timing parameter  $U$  on loss rate

In addition to the logical topology design and routing aspects of the algorithm, link up and downgrade are now also part of the MTE process. Link up/downgrade will add or remove lightpaths to IP links on-demand. This process itself is also timed with a parameter  $U$  (same unit as  $T$ ). For example, a practical implementation would measure traffic on a link during each  $U$  time steps, then adjust link capacity. We will briefly examine the impact of this parameter on traffic loss, using similar simulations, however with higher volumes of bulk and additional traffic ( $\lambda = 0.2$  and  $\mu = 0.01$ ), corresponding with the higher amount of available bandwidth.

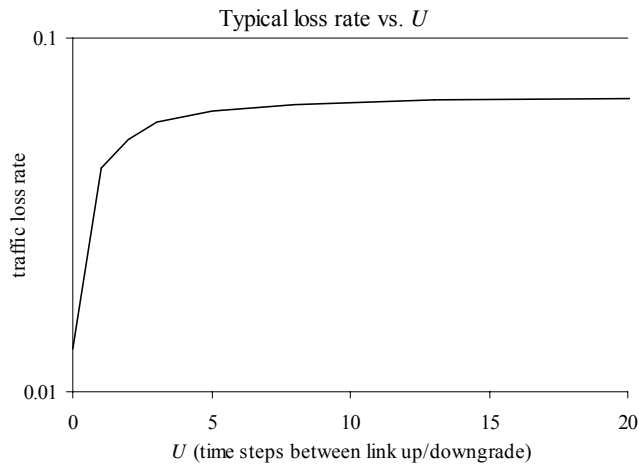


Figure 9. Typical loss rate in function of  $U$

On Fig. 8 we see how this link bandwidth adjustment process impacts traffic loss (again also vs. bulk traffic volume). The loss rates for  $U=0$  represent an ideal case where all up or downgrades are performed immediately. For increasing  $U$  (increasing ‘lag’ on the bandwidth adjustment actions), we see loss shoot up from a baseline 1% up to 10%. Loss rates seem quite sensitive towards this parameter, as can

be seen on Fig. 9 as well, where loss rate is charted vs.  $U$  for typical moderate bulk traffic volumes.

Luckily however, the technical complexity of link bandwidth adjustment is somewhat limited compared to routing and logical topology design, which happen on a network-wide scale. As [5] has shown, link bandwidth adjustment using optical layer switching is certainly possible on short timescales.

## V. CONCLUSION

The multilayer traffic engineering strategy discussed can perform logical topology design and routing of traffic flows for an IP-over-Optical network scenario. Some of the problems that stem from trade-offs between optical layer capacity usage and IP layer performance may be alleviated by using more aggressive timing in this proactive MTE strategy. Conversely, relaxing the amount of optimization (and specialization) of the resulting logical topologies may allow an operator to keep acceptable performance where technical issues limit the frequency of adaptations – and therefore the speed at which the MTE algorithm can react, etc.

The concatenation of multiple parallel lightpaths into single IP links can be used to avoid dense meshing in the IP layer, and to keep the MTE algorithm in a more point-to-point grooming regime, even for larger traffic volumes.

## ACKNOWLEDGMENT

Bart Puype and Didier Colle thank the IWT for its financial support for their PhD and postdoctoral grants respectively. The work was partly funded by the European Commission through the projects IST-NOBEL and IST-ePhoton/ONE; by the Flemish Government through the projects IWT-GBOU ONNA and FWO G.0315.04.

## REFERENCES

- [1] B. Puype et al. "Multi-layer Traffic Engineering in Data-centric Optical Networks, Illustration of concepts and benefits," in Proc. ONDM 2003, Budapest (2003), pp. 221-226.
- [2] B. Puype et al. "Optical cost metrics in Multi-layer Traffic Engineering for IP-over-Optical networks," in Proc. ICTON 2004, Wroclaw (2004), vol. 1; pp. 75-80.
- [3] Q. Yan et al., "Influence of the observation window size on the performance of multi-layer traffic engineering," in Proc. ITCOM, Orlando (2003), Proc. of SPIE Vol. 5247; pp. 203-214.
- [4] E. Van Breusegem et al., "Evaluation of ORION in predimensioned networks," in Proc. ITC19, Vol. 6b, Beijing (2005); pp. 1265-1274.
- [5] K. Sato et al., "GMPLS-Based Photonic Multilayer Router (Hikari Router) Architecture: An Overview of Traffic Engineering and Signaling Technology," IEEE Communications Magazine, Vol. 40, No. 3, pp. 96-101, March 2002.