



Performance Evaluation of Multi-Layer Traffic Engineering Enabled IP-over-ION Networks

Qiang Yan, Didier Colle*, Sophie De Maesschalck, Bart Puype, Ilse Lievens, Mario Pickavet, Piet Demeester

Department of Information Technology, Ghent University - IMEC, Sint-Pietersnieuwstraat 41, 9000 Ghent, Belgium
E-mail: {qiang.yan, didier.colle, sophie.demaesschalck, bart.puype, ilse.lievens, mario.pickavet.piet.demeester}@intec.UGent.be

Received February 26, 2004; Revised September 24, 2004; Accepted September 27, 2004

Abstract. Recently, network operators started implementing traffic engineering (TE) techniques in their network. These TE techniques typically involve a single layer (for example, the IP/multi-protocol label switching (MPLS) layer). Although single-layer TE (STE) can improve the network performance (e.g., throughput, quality of service (QoS)), this improvement is bounded by the available capacity in that network layer. The evolution towards intelligent optical networks (IONs) allows further increasing the improvements achievable by the TE techniques, by involving more than one layer in the TE actions. Multi-layer TE (MTE) occupies network resources in a smart way and optimizes the QoS since it dynamically reconfigures the logical topology in the upper layer by properly updating the optical connections in the underlying optical layer. However, the performance of the network is impacted by the configuration scheme adopted by MTE. Therefore, in this paper, we focus on analyzing the influence of the MTE configuration scheme on the MTE behavior, and evaluate the network performance by studying simulation results obtained from a realistic IP-over-ION network.

Keywords: multi-layer traffic engineering, IP-over-intelligent optical network, dynamic grooming

1 Introduction

Today, optical fibers equipped with wavelength division multiplexing (WDM) technology provide tremendous amounts of capacity in a cost-efficient way. This means that in most cases the transmission bandwidth is no longer the limiting factor in IP-over-Optical networks. By introducing optical cross-connects (OXC) in the optical transport network (OTN), the OTN is able to set-up and tear-down lightpaths by properly configuring the OXC. Each lightpath occupies one or more wavelength channels in the fibers along its path. Each wavelength channel typically has a capacity of 2.5 or 10 Gbps, which is considerably high compared to the rate of a single IP traffic flow. Therefore, aggregating several low-rate traffic flows into a single lightpath, called traffic grooming, is an efficient and much practiced way to use the tremendous capacity offered by an OTN.

There exist a lot of strategies to groom the traffic: for example, in the end-to-end grooming and link-by-link grooming strategies, respectively, the lowest amount and the highest amount of traffic is groomed together in a lightpath. As illustrated in Fig. 1, end-to-end grooming means that the traffic between a pair of IP routers is aggregated into a dedicated optical connection (i.e., lightpath) from source to destination node in the underlying layer, without any intermediate transfer back to the IP layer. In other words, the logical IP topology contains for each IP traffic flow a direct end-to-end logical IP link. End-to-end grooming can reduce the transit traffic demand in the IP routers, but is quite expensive in terms of wavelength channels. Vice versa, with link-by-link grooming, intermediate transfers to the IP layer are made in each intermediate node. In other words, the logical IP topology is identical to the physical (optical) topology. The link-by-link grooming method

*Corresponding author.

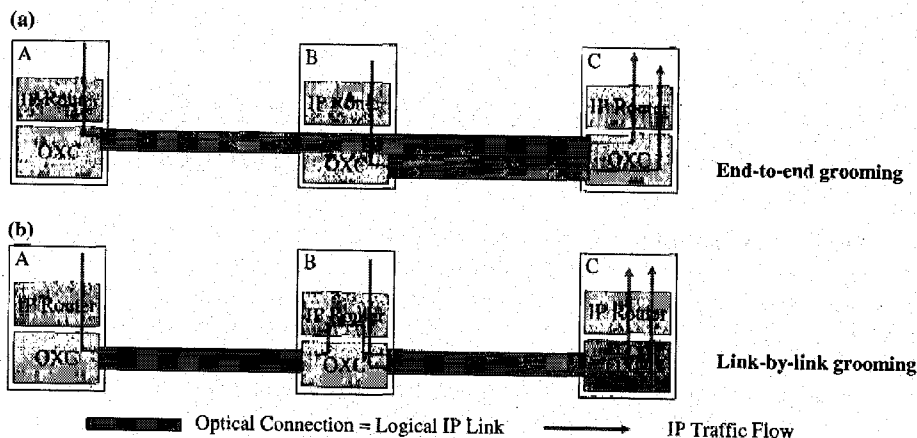


Fig. 1. Traffic grooming.

increases the transit traffic demand in the IP routers, but increases the filling grade of the wavelength channels. Link-by-link and end-to-end grooming are the two extreme options when it comes to grooming. Different solutions that fall in between both extremes have been discussed in literature. They may envisage different goals: minimizing the overall network cost, maximizing the throughput, etc. Refs. [1,2] give an excellent overview of traffic grooming in WDM networks. Since the grooming problem has mainly been studied for ring networks, the focus of [1,2] is on SONET/SDH and WDM ring networks, but also the grooming problem in meshed IP-over-WDM networks is described. These papers explain how grooming can reduce the network cost and increase the utilization grade of the network connections. The work described in [3,4] is specifically aimed at finding a good solution for the grooming problem in mesh optical networks. In [3] the objective is to improve the network throughput by deploying grooming and an ILP formulation is used to solve the problem. In [4] the grooming problem in an IP/multi-protocol label switching (MPLS)-over-optical network is tackled in order to find the lowest overall network design cost using a heuristic approach. Both methods demonstrate the value of grooming in optimizing the network design.

Up till now, the OTN can only be statically configured by the operator through the network management system. The IP traffic flows are groomed into permanent lightpaths. Satisfying a request to set-up or tear-down a lightpath needs a

lot of manual intervention, resulting in a long response time. Therefore, considering that Internet traffic shows a dramatic increase and a continuously changing pattern, this static provisioning of capacity would require an important over-provisioning. Being able to dynamically reconfigure the logical network would significantly improve the capacity efficiency.

To cope with the dynamic traffic pattern, current research is investigating the opportunity to introduce more intelligence and autonomy into the OTN in the form of a distributed control plane. This type of optical network is denoted as intelligent optical network (ION), also termed automatic switched optical network (ASON) by ITU [5]. IONs allow clients to trigger the set-up and tear-down of a connection by sending a signal into the transport network through the user-network interface (UNI) [6]. Connections established in this manner are called switched connections instead of (soft-)permanent connections. In this way, any manual intervention of the network operator is omitted, which results in an increased network flexibility.

In an IP-over-ION network, the IP layer network is considered as the logical network whose nodes are connected by the optical connections offered by the ION. This means that each (logical) link in the IP network maps to an optical connection in the optical network. By monitoring the traffic load on the IP links, nodes in the IP network can request the underlying ION to set-up or tear-down one or more optical connections by signaling through the UNI.

Within the IP layer, MPLS and constraint-based routing are often used to set up a label switched path (LSP) to satisfy the client's agreed quality of service (QoS) [7–9]. This enhances the performance of traditional IP layer traffic engineering (TE). Based on traffic measurements, IP traffic flows can be routed dynamically, improving the QoS as described in [10,11] or anticipating congestion problems as described in [12]. Furthermore, in case of a capacity shortage in the IP network layer, the TE engine could request the optical network to provide some more capacity. As the underlying network layer is now also involved in the TE, the logical network capacity can be adapted to the traffic load at any time. This is called multi-layer TE (MTE). By dynamically reconfiguring the logical network (topology), the traffic grooming also becomes dynamic, which is called dynamic grooming. For example, when in Fig. 1 both traffic flows become too large to be groomed together on the logical link B–C, the MTE will decide to replace the logical IP link between routers A and B by a direct logical link between A and C (which requires two instead of one wavelength channels on the physical link B–C), thus evolving from link-by-link towards end-to-end grooming.

A recent MTE implementation is the Hikari router [13]. To realize MTE, the Hikari router uses an open shortest path first (OSPF) [14] extension, which advertises both the total number of wavelengths and the number of unused wavelengths, and a RSVP-TE (Resource reSerVation Protocol – Traffic Engineering) [15] extension, which minimizes the number of wavelength conversions needed. In addition, paper [13] proposes a heuristics-based multi-layer topology design scheme that uses IP traffic measurements in a generalized multi-protocol label switch (GMPLS). Ref. [16] proposes a GMPLS-based MTE scheme that combines periodical offline global reoptimization with interim online dynamic routing in an IP-over-WDM multi-layer scenario using label switching routers (LSRs) and OXCs.

In this paper, we investigate different MTE strategies in order to find their influence on the offered QoS (in terms of packet loss, etc.) and the used resources (in terms of wavelength channels) while dynamically reconfiguring the logical network on top of the ION.

The paper is organized as follows. First, we will introduce the three phases involved in MTE in Section 2 and discuss from a theoretic viewpoint the influence of the applied method on the QoS. Also, the MTE strategy adopted in the simulations will be described. Then, the simple case study of Section 3 confirms the influence of the configuration of the MTE (more in particular the configuration of the traffic observation window (TOW) size, see further) on the QoS. Afterwards, in Section 4, a detailed investigation of the influence of different parameters on the performance of the proposed MTE strategy is carried out based on simulation results obtained from an Italian nationwide OTN. Finally, in Section 5, we will recapitulate the major findings discussed in this paper.

2 Multi-Layer Traffic Engineering

A too simple and straightforward routing strategy is typically unable to fairly balance the traffic load over the network, and can thus lead to congestion in part of the network as well as decrease the QoS, even though sufficient capacity is available. TE can be introduced in order to route the traffic more evenly over the network while taking into account a certain level of QoS.

Regular TE is only working at one layer; therefore we use the term single-layer TE (STE). E.g., in an IP-over-Optical network, the upper layer (IP layer) generally uses MPLS and Constraint Based Routing to achieve the objective of guaranteeing the user a satisfactory QoS [7–9]. As the upper part of Fig. 2 illustrates, TE in the IP-MPLS network routes a new connection with a demand equal to 60% of the capacity of an IP-MPLS link connection through the lower loaded links (a–c–b–d–e) instead of via the least-hop path (a–c–e) because otherwise the link c–e would get congested (200% is available while $160\% + 60\% = 220\%$ would be needed).

In contrast with the objective of TE in the IP-MPLS layer, the goal of TE in the optical layer is to reduce the blocking of connections due to a lack of wavelength channels. In other words, the objective of TE in an optical network is to achieve a lower blocking ratio by efficiently using the available network resources. As the middle part of Fig. 2 shows, TE in the optical layer routes the

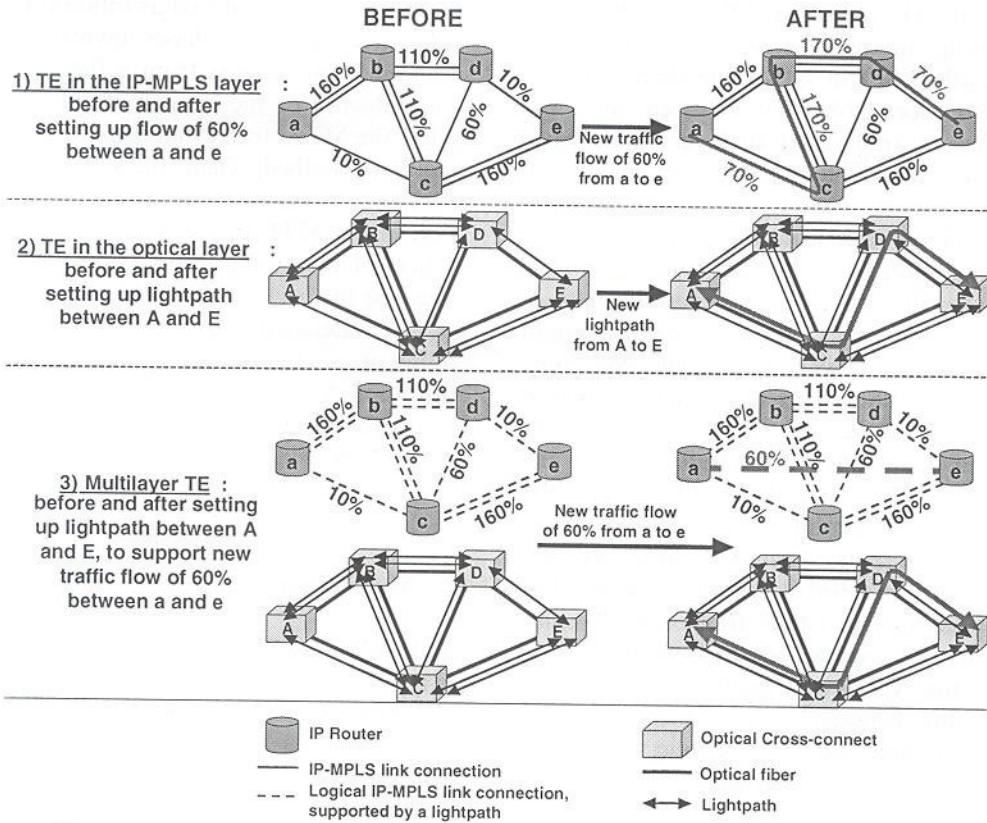


Fig. 2. Concept of STE and MTE (100% refers to the capacity of an IP-MPLS link connection, corresponding to a single optical connection).

lightpath along the path going through the least occupied fibers (a-c-d-e) instead of the shortest path (a-c-e) in order to avoid blocking due to a lack of wavelength channels (for example consider that each fiber only carries two wavelengths channels).

However, upper layer TE (i.e., IP-MPLS layer TE) can only realize limited optimization. As shown in the upper part of Fig. 2, although the path goes through the lower loaded links (a-c-b-d-e), the transit traffic demand through nodes B and D is increased substantially. Thus the packets may have to be queued in the buffer before being transported, which implies an incremental end-to-end delay. In order to avoid the deterioration of the QoS due to the incremental buffer occupancy, a solution is to reconfigure the logical topology in the IP-MPLS layer by updating the optical connections in the optical layer (see bottom of Fig. 2). Such a TE strategy, involving multiple layers, is

termed multi-layer TE (MTE). MTE routes the new traffic demand through a new logical link (a-c), which neither increases the traffic demand in the existing links nor adds extra transit traffic in the other nodes.

In general, MTE strategies are split up into two classes: proactive strategies [17] and reactive strategies [18]. A proactive strategy is a strategy that continuously updates the logical network in order to keep the network away from undesirable situations. However, a proactive strategy might affect the network stability because of frequent updates. On the other hand, a reactive strategy is only triggered when a network congestion or QoS degradation is detected. Compared to a proactive strategy, a reactive strategy limits the number of reconfigurations. In [17,18], the authors explain that all MTE strategies involve three phases: triggering (see Section 2.1), decision making (see Section 2.2) and realization (see Section 2.3). In

the remainder of this section, we will propose an efficient reactive MTE strategy in detail.

2.1 MTE Triggering Phase

No matter what strategy is used, the objective is to keep the network at a desirable situation in terms of QoS and stability (i.e., few logical network reconfigurations) by dynamically reconfiguring the logical network. In order to reconfigure the logical network in the right way at the right time, MTE must know the recent traffic load situation accurately. For this purpose, a practical solution is to count the number of packets injected into a logical link in the upper layer during a fixed time segment, which is called the TOW. This real time traffic demand calculation method is illustrated in Equation (1). D is the traffic demand through a certain link; n is the number of traffic flows through this link; d_i is the demand of a certain traffic flow in this link; S_i is the average packet size in bytes; N_i presents the number of packets that a certain traffic flow injected into the link during the TOW. T_{TOW} is the size of the TOW in seconds.

$$D = \sum_{i=1}^n d_i = \sum_{i=1}^n \frac{S_i \times N_i}{T_{TOW}} \quad (1)$$

In the proposed reactive MTE strategy, the evaluated traffic load is compared to two pre-defined thresholds: Th_{high} and Th_{low} . The former threshold

indicates the situation of traffic congestion while the latter indicates an underloaded situation. MTE is triggered to add a new link when the traffic load is above Th_{high} and to remove a link when the load is below Th_{low} .

From Equation (1), it is clear that different reactions to a certain traffic pattern may be triggered when different TOW lengths are used. Let us examine the example illustrated in Fig. 3. Suppose that the dashed lines present the detected traffic demand under different TOW lengths. Case (a) detects every traffic overload (demand above Th_{high}) and underload situation (demand below Th_{low}) in the traffic pattern by using a small TOW, which is sensitive to the traffic variation. Such small TOW will trigger the MTE to add a new logical link to the upper layer network at time 2 and 4 s to handle the detected traffic burst, as well as to remove a logical link at time 5 s due to the underloaded network situation. In case (a), the logical network is reconfigured each time a logical link is overloaded or underloaded. In case (b), the TOW length is increased to 2.5 s. As shown in Fig. 3b, the MTE is only triggered at the end of the first TOW (2.5 s) although during both TOWs traffic bursts occur (traffic demand above Th_{high}). Compared with the small TOW used in case (a), the TOW used in case (b) may avoid unnecessary actions (like reacting to a very short overloaded or underloaded situation), which implies a reduction of the number of logical network reconfigurations.

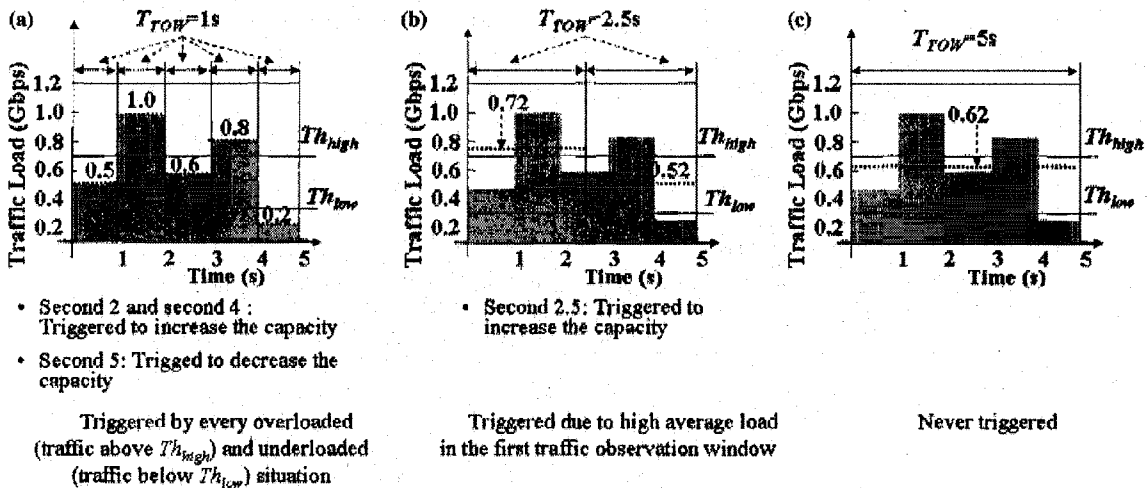


Fig. 3. Traffic monitoring and MTE triggering.

Fig. 3c involves an even larger TOW. Since the average load over a long period hides the detail of traffic variations such as overloading or underloading, a large TOW may not be able to detect the traffic congestion timely, and will result in no or an inappropriate reaction of the MTE. In case (c), MTE will not reconfigure the logical network to be suited for the traffic pattern because the detected traffic load in the TOW lies between Th_{high} and Th_{low} .

Summarizing, it is clear that different TOW lengths will result in different reactions (different logical network reconfigurations), and thus impact the QoS (e.g., packet loss ratio) differently. A simple case will be studied in Section 3 to present the influence of the TOW length on the network performance.

2.2 MTE Decision Making Phase

The work flow of a reactive MTE strategy is illustrated in Fig. 4. Once an undesirable situation has been detected, MTE starts to plan the upper layer network reconfiguration scenario. This phase is termed decision making phase. To cope with different traffic load situations, this process involves three branches, one to cope with traffic underload, one for a desirable load traffic and one for traffic overload.

When a certain logical link is underloaded (traffic demand D is lower than Th_{low}), keeping it

in the logical network may be not valuable. Thus, MTE removes this link by tearing down the corresponding optical connection in the underlying optical layer. However, just simply tearing down a logical link each time MTE detects an underloaded situation would not be desirable, since the action of tearing down a logical link may be triggered by a very short underloaded period and the removed logical link may be reintroduced soon (see also case study in Section 3). In order to improve the network stability and reduce the packet loss caused by the tear-down action, such unnecessary actions should be avoided. Therefore, we introduce a counter (*underloaded_counter*) for each logical link into the proposed MTE strategy to count the number of consecutive detected underloaded situations. At the end of each TOW, the counter is either incremented in case an underloaded situation is detected, or reset to zero otherwise. The logical link is considered as really underloaded and will actually be torn down when the counter equals n . This implies a reduction of the number of reconfigurations. After a logical link is torn down, the traffic flows passing through the removed logical link are rerouted. Since releasing an optical connection in the optical layer is a basic function for an optical control plane, and the IP routing protocol (e.g., OSPF) can automatically update the routing table in each router

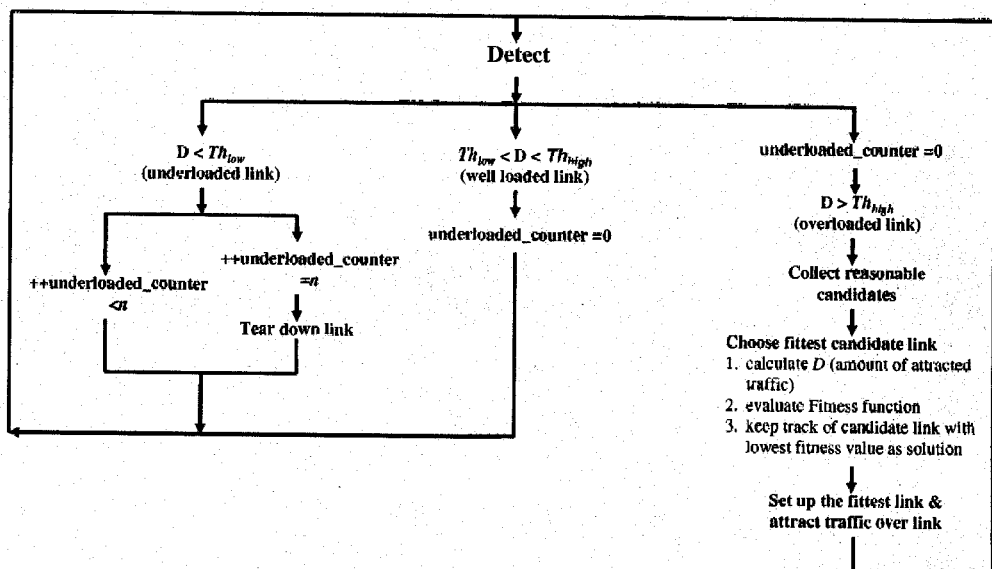


Fig. 4. Reactive MTE strategy work flow.

when the state of an IP link is changed, implementing the reaction to the underloaded situation is straightforward. Thus, the following discussion will put more emphasis on implementing the reaction to congestion (overloaded situation).

If the logical link is considered to be in the well loaded situation ($Th_{low} < D < Th_{high}$), no action will be performed except resetting the *underloaded_counter* to zero.

When a link is considered congested (traffic demand D is higher than Th_{high}), this means the actual logical network capacity is not sufficient for the traffic load. Unlike the action coping with traffic underload, we do not propose an overload counter since it would introduce delay into reacting to the overload and consequently result in an increased packet loss. Therefore, MTE is immediately triggered to add a new link to the logical network or to increase the capacity of an existing link when an overload situation is detected. The problem arising here is how to locate the fittest candidate link from many possibilities. A solution is to use the heuristic algorithm of Section 2.2.1 to collect reasonable candidates for the new link, and to adopt the fitness function described in Section 2.2.2 to evaluate the fitness of each candidate link.

2.2.1 Heuristic for Determining The Candidate Links

As we know, there are plenty of possibilities to reconfigure the upper layer network when congestion occurs. Which candidate link is the best one to be set up in the logical network? We propose to investigate the traffic flows through the congested link and then collect the reasonable candidate links obeying the following rules:

- The traffic flow over a congested link, whose demand is higher than Th_{high} (per optical

wavelength channel), should be transported through a direct logical link from source to destination. This is because the larger demand could cause congestion in several links and nodes along its path.

- Let us consider the case shown in Fig. 5. Traffic flows f_1, f_2 and f_3 pass through the congested link B-C. There are common segments in the paths of the flows, i.e., traffic flows f_1 and f_2 both pass through the links A-B and B-C. Since adding a link between the end points of the common part can remove one or more flows from the congested link, it can reduce the traffic load in the congested link efficiently. Therefore, picking out the common path between the flows through a congested link is an efficient way to find reasonable candidate positions for a new link.

In conclusion, each candidate link to be set up in the network corresponds to a path that is a common route for one or more traffic flows through the congested link in the current logical network.

2.2.2 Fitness Evaluation of The Reasonable Candidate Links

Before discussing the objective function used to evaluate the fitness of a candidate link, we have to explain an important assumption we made here, namely that the logical network uses a link-state routing protocol, i.e., the OSPF routing protocol. Any state change in a logical link will be detected by the source of the logical link and notified to the others by sending the link-state advertisement messages. This means that the establishment or tear-down of a logical link will be automatically detected and notified to the others by the source node of that logical link. Then, the link-state database and routing table in each node in the

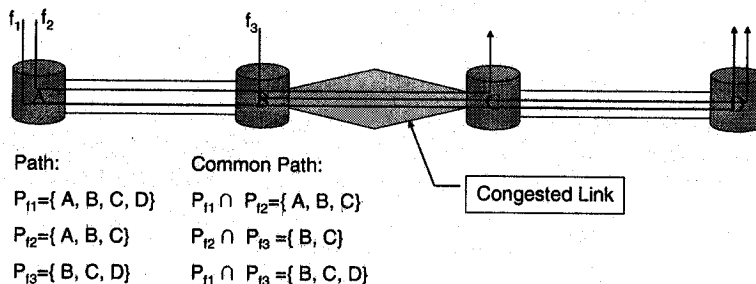


Fig. 5. Common path consideration.

logical network are updated. Since OSPF is based on the Dijkstra algorithm, links with a lower cost are preferred. Therefore, by assigning to the candidate link a cost lower than the cost of the common path replaced by the selected candidate, the new link will automatically be involved in the new routes of the traffic flows. In other words, the traffic flows through the replaced common path will be automatically rerouted through the new link. On the other hand, assigning the new link a cost that is large enough prevents that other traffic flows are rerouted.

How can we determine the fittest candidate out of the list of candidate links to add? In order to answer this question, we consider the following objectives for reconfiguration.

1. After the reconfiguration the currently congested link should not be congested anymore. By adding a new link in the logical network, part of the traffic demand through the congested link is swapped to the new link: this should resolve the congestion on the currently overloaded link. This rule aims at guaranteeing that the spare capacity remaining in each logical link is enough to avoid packet loss during a short traffic burst.
2. After the reconfiguration, the traffic load in both the new link and the currently congested link should be fair and close to the middle between Th_{low} and Th_{high} . A motivation is that this rule should result in a fair traffic load in all logical links in the network.
3. After the reconfiguration, the transit traffic demand in some of the nodes should be reduced. When a new link is added to the logical network, the hop-length of the paths for some traffic flows is reduced. This implies a reduction of the transit traffic

demand as well as a reduction of the buffer occupancy in some upper layer nodes. Consequently, the end-to-end delay of a packet will be shortened and the probability to get lost will drop.

Suppose the capacity of a certain link i is denoted as W_i . There are two thresholds, $Th_{high,i}$ and $Th_{low,i}$ (denoted as a percentage of the link capacity), defined as the MTE triggers. In order to keep fair loading in all logical links of the network and to avoid traffic overloading, the optimal traffic demand passing through link i should be $D_{mid,i}$, which is defined in Equation (2). Consequently, each link has some spare capacity to avoid an overload situation during a short traffic burst.

$$D_{mid,i} = \frac{1}{2} (Th_{high,i} + Th_{low,i}) * W_i \quad (2)$$

As illustrated in Fig. 6, assuming the demand through a congested link is D_{cong} , a fraction Δd of the demand will be moved from the congested link to the new link after the reconfiguration. Thus, after the reconfiguration, the demand left in the previously congested link (D_{final_cong}) is $D_{cong} - \Delta d$. After the reconfiguration, we would like to have a remaining demand in the congested link approaching D_{mid_cong} , and a rerouted demand approaching the preferred demand of the new link D_{mid_new} . Therefore, Equation (3) defines the optimal demand Δd_{opt} to be deviated from the congested link, as the average of the expected demand to be deviated from the congested link ($D_{cong} - D_{mid_cong}$) and the ideal demand attracted by a new candidate link (D_{mid_new}). To cope with the third criterion, longer candidate links (in number of hops) should be preferred in order to reduce the transit demand in more nodes.

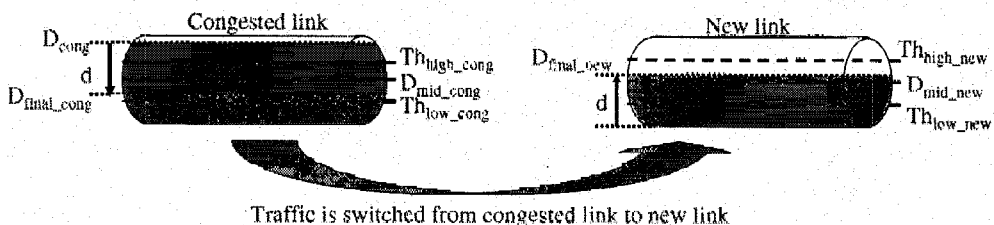


Fig. 6. Traffic distribution after reconfiguration.

$$\Delta d_{opt} = \frac{1}{2}(D_{cong} - D_{mid,cong} + D_{mid,new}) \quad (3)$$

$$\text{Min } F_j(\Delta d_j, \#Hops) = \frac{[\Delta d_j - \Delta d_{opt}]^2}{\Delta d_j(\#Hops - 1)} \quad (4)$$

Therefore, we propose the fitness function of Equation (4). The upper part of this equation indicates the difference between the real demand attracted by the candidate link and the optimal demand as calculated in Equation (3), which should result in a fair traffic loading after the reconfiguration. Candidate links resulting in a smaller difference are preferred here. The lower part is the product of the traffic load that would be attracted over the candidate link and the length of that candidate link, corresponding to the reduction of the transit traffic demand in the intermediate nodes in the logical network. The candidate with the smallest fitness value will be selected to be set up.

2.3 MTE Realization Phase

Once the decision has been made to remove or to add a logical link in the logical network, it should

be carried out immediately. In the MTE-enabled network illustrated in Fig. 7, the logical link b-c is congested, as it is overloaded by the traffic flows a → d and b → c. In order to solve the congestion, MTE is triggered to reconfigure the upper layer logical network. Supposing the goal of the reconfiguration is to establish a logical link between router a and d, this actually means that a new optical connection has to be requested between OXCs A and D in the underlying physical network. Initially, as a decision maker, router b sends a suggestion to router a, which is attached to the source of the new optical connection (OXC A) in the optical layer, to start the process of establishing an optical connection.

Several practical solutions exist for signaling the set-up or tear-down of an optical connection; e.g., OIF user-network interface (UNI) allows the upper layer routers to send signals into the underlying physical network to request an optical connection. In an IP-over-Optical network, GMPLS adopts the RSVP-TE or CR-LDP (constraint-based routing-label distribution protocol) [19] protocol for setting up or tearing down an optical connection. Suppose the RSVP-TE messages are used to set up an optical connection. Once router a receives the suggestion, it asks OXC A to send out the Path message to OXC D. The Path message requests the

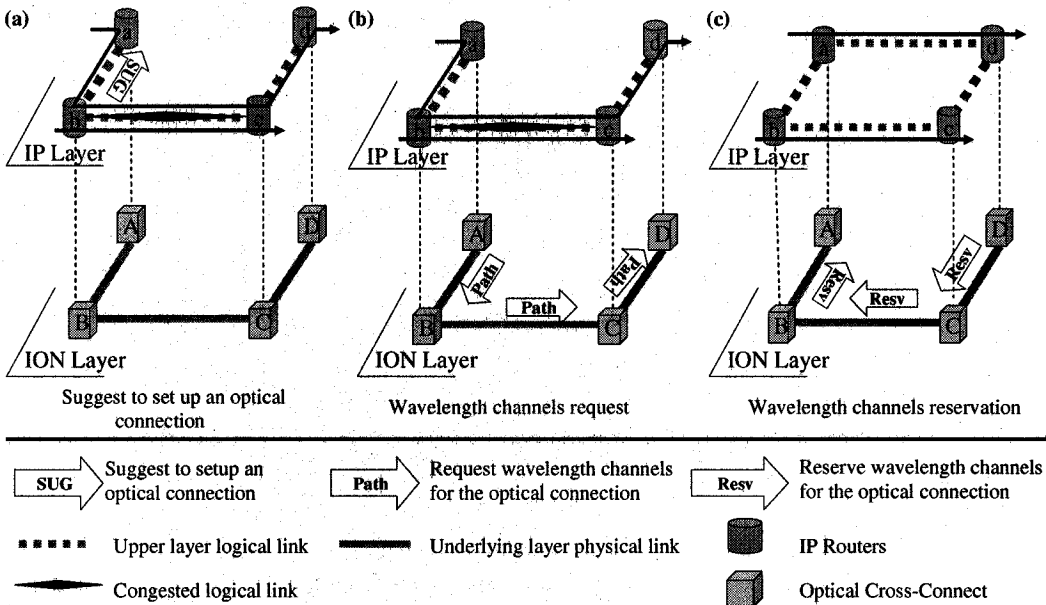


Fig. 7. Reconfiguration realization (using RSVP-TE messages [15]).

wavelength channels for the connection along its path. After OXC D receives the *Path* message, it sends back a *Resv* message to reserve the wavelength channels. As soon as OXC A receives the *Resv*, it activates the optical connection as well as notifies a new logical link to the upper layer network. Finally, the traffic $a \rightarrow d$ is attracted over the new link.

However, the request to set up an optical connection may be blocked due to a lack of wavelength channels in a fiber along the path. As shown in Fig. 8, since some fibers are heavily occupied, simple shortest path routing does not make sense when a new connection needs two wavelength channels, in this example. In order to establish this new connection, either routing the connection through a longer path, along which each fiber offers enough wavelengths, or routing the connection through several shorter paths, each one offering part of the capacity, makes the new connection realizable. The former method is called single path approach (SPA), and the latter one is termed multiple path approach (MPA). In the situation of Fig. 8, SPA could not add a connection with a capacity equal to three wavelengths under the current situation, but MPA still works. Therefore, in order to efficiently reduce blocking in the physical network, the enhanced MPA routing method is adopted. The goal of this method is to avoid heavily occupied wavelengths in part of the network by using a simple and straightforward scheme.

3. Case Study

In the previous section, we introduced MTE in detail. The objective of MTE is to reconfigure the logical network so that a desirable QoS is kept over time. However, it may not work properly if it is not configured correctly. E.g., different TOW lengths result in different reconfigurations of the logical network as well as different grades of QoS offered by the network. The cases presented in this Section illustrate and confirm the benefits of MTE as well as study the influence of the TOW length.

As mentioned before, MTE could react differently to the same traffic pattern by assigning a different length to the TOW. This phenomenon is confirmed by the simple case illustrated in Fig. 9.

Suppose we have a small network as shown in Fig. 9. Each link has a delay of 1ms. Two traffic flows are transported over this network, one from node A to C (f_A), and the other from node B to C (f_B). Both follow the shown traffic pattern and start at the same time. Each wavelength channel has a capacity of 2.5 Gbps. Since each logical link occupies at least one wavelength channel, the minimal capacity of a logical link is 2.5 Gbps. The initial upper layer logical topology equals the underlying physical topology. This means that the upper layer traffic flows are link-by-link groomed into the underlying layer at the beginning. Each node is MTE-enabled. The traffic evaluation thresholds are configured as follows: T_{high} is set

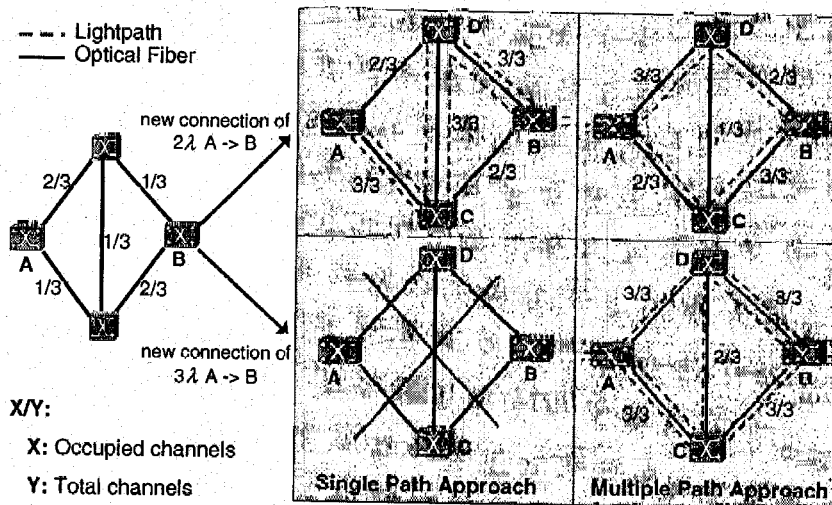


Fig. 8. Single path approach (SPA) vs. multiple path approach (MPA).

to 70% (=1.75 Gbps) of the total logical link capacity; Th_{low} is set to 30% (=0.75 Gbps) of the total logical link capacity. The network performance in terms of packet loss is evaluated every 100 ms. Finally, a logical link is torn down when it has been detected to be underloaded in three consecutive TOWs (i.e., $n=3$). The upper charts in Fig. 10 illustrate the packet loss ratio when different TOW lengths are used while the lower charts illustrate the detected utilization grade of the logical links B-C and A-C over different TOW lengths.

Considering the packet loss ratio (PLR) shown in Fig. 10, the size of the TOW actually impacts the QoS offered by the network. Fig. 10a presents the packet loss ratio situation when a small TOW of 0.5 s is used. The utilization grade curve of logical link A-C (see Fig. 10b) is interrupted in those periods that logical link A-C is not present in the logical network. When the traffic burst highlighted by circle X in Fig. 9 comes, MTE is triggered to add the new link A-C to reduce the congestion in link B-C. However, link A-C is not stable. This is actually because the duration of the traffic underloading, marked by circle Y in Fig. 9, is longer than the TOW length. During the underloaded period (in circle Y) as the link A-C is detected to be underloaded in three consecutive

TOWs, MTE is triggered to release this link. Afterwards, the link A-C is set up again since MTE is triggered by the following burst right after the underloaded situation. Thus, link A-C is established and released repeatedly over time. Each time link A-C is torn down, the packets transported along A-C are dropped, which causes 1% of the packets counted in a period of 100 ms to get lost. Each reconfiguration causes the traffic flow f_A to be swapped between path A-C and path A-B-C. Therefore several short period peaks appear along the curve of the packet loss ratio in Fig. 10a at the time the link A-C does not exist. This case shows that MTE using a small TOW is able to reduce part of the packet loss, but that the logical network is too frequently reconfigured due to unnecessarily tearing down the logical link A-C.

Chart b in Fig. 10 presents a desirable situation. In this case, when the first burst (marked by circle X in Fig. 9) arrives, MTE is triggered to add the new link A-C to the logical network due to the higher average load during the TOW. Since the MTE action is performed at the end of a TOW, the larger the TOW is, the slower the MTE reacts to the traffic increment. Therefore, we observe that the period of serious packet loss in Fig. 10b is larger than in Fig. 10a. After the reconfiguration,

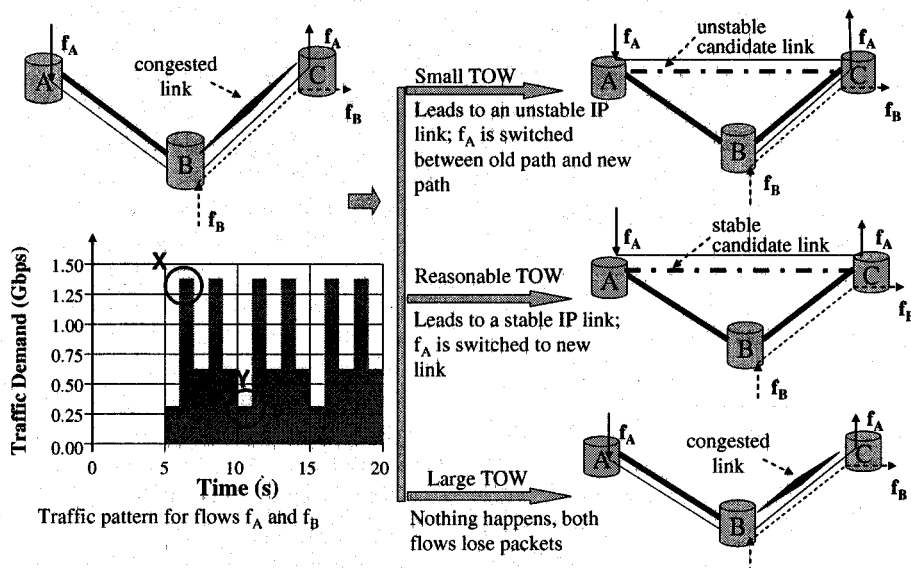


Fig. 9. Influence of the TOW length.

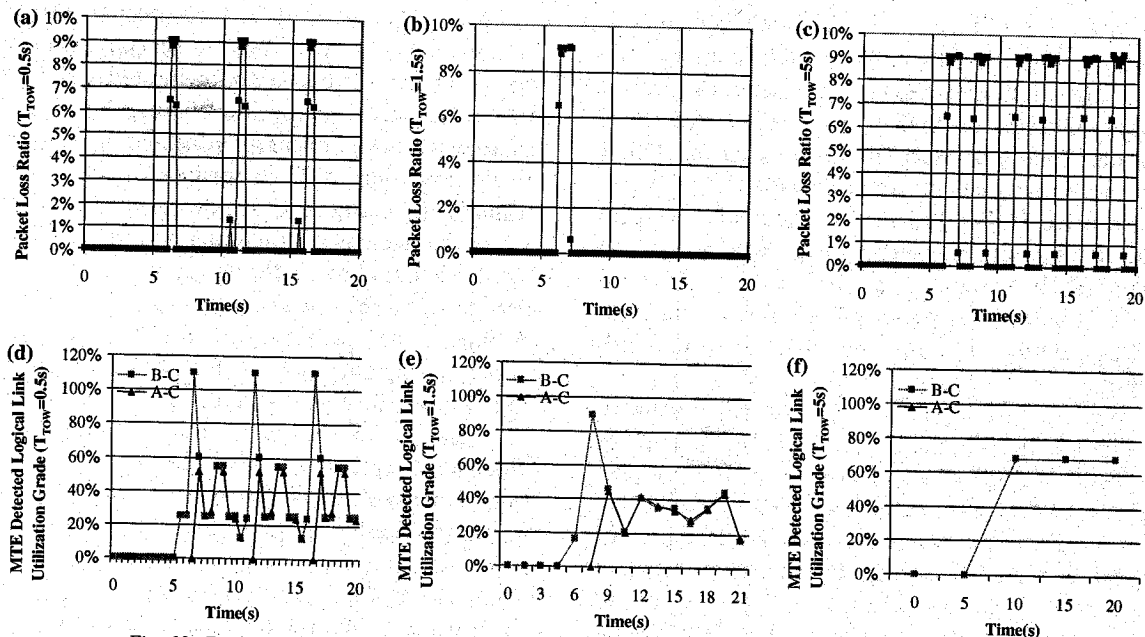


Fig. 10. Packet loss ratio (PLR) and detected link utilization grade over different TOW lengths.

the link A-C remains in place since the average load within a TOW does not get below Th_{low} in three consecutive TOWs. This case illustrates the challenge of keeping the QoS at a desirable level over time by assigning a suitable length to the TOW used by the MTE scheme.

The last case in Fig. 10 illustrates the possibility that the MTE loses control of reacting to traffic variations. This is actually caused by the average demand evaluation character of a TOW. Since the average traffic demand over a large TOW is below Th_{high} , no action is performed by the MTE. Consequently, the logical network cannot be duly reconfigured to cope with the traffic burst, which results in the peaks in the packet loss ratio charts in Fig. 10c. Meanwhile, the link utilization curve of link A-C does not exist in Fig. 10f.

4. Simulation Study

Previous sections discussed the benefits of MTE and indicated the influence of its configuration scheme on the QoS offered by an MTE-enabled network. In this section, we will use MTE in a realistic nation-wide IP-over-Optical network to confirm the analysis presented before. The simulation platform used is Ns2 [19].

4.1 Simulation Scenario

All the results discussed in this section are obtained from simulations on a national IP-over-Optical network in Italy. The topology of the Italian network is illustrated in Fig. 11. It includes 14 MTE capable nodes and 29 bi-directional optical fibers. On each link, both directions contain the same number of channels. The capacity is denoted as the capacity in one direction, i.e., when the capacity of a link is denoted as 5 wavelength channels, it means that the corresponding fiber pair carries in each direction 5 wavelength channels. Each wavelength channel has a capacity of 2.5 Gbps. Each node is MTE-enabled and could be considered as an integrated box with an IP router and an OXC. Maximum 10 packets of 64 KB each are allowed to wait in the output buffer in each network interface of an IP router.

The traffic between a pair of nodes is modeled as random self-similar traffic by using an on-off traffic model, with an on period of 40 ms on average, and an off period of on average 10 ms. The lengths of the on and off periods are Pareto distributed. Thus, the integrated traffic demand in a logical link has a self-similar character [21]. The traffic pattern is asymmetric.

Actually, we use two scenarios in the following simulation. One is the constant traffic injection

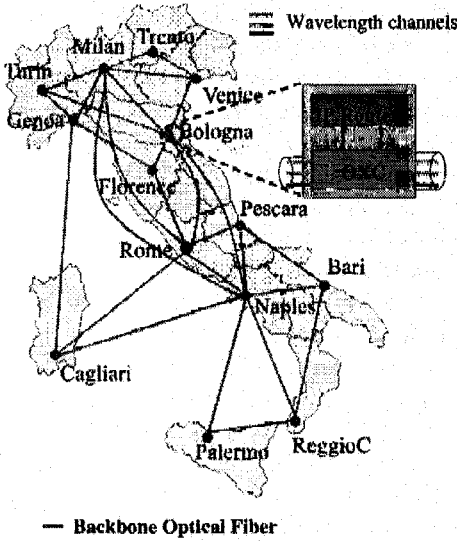
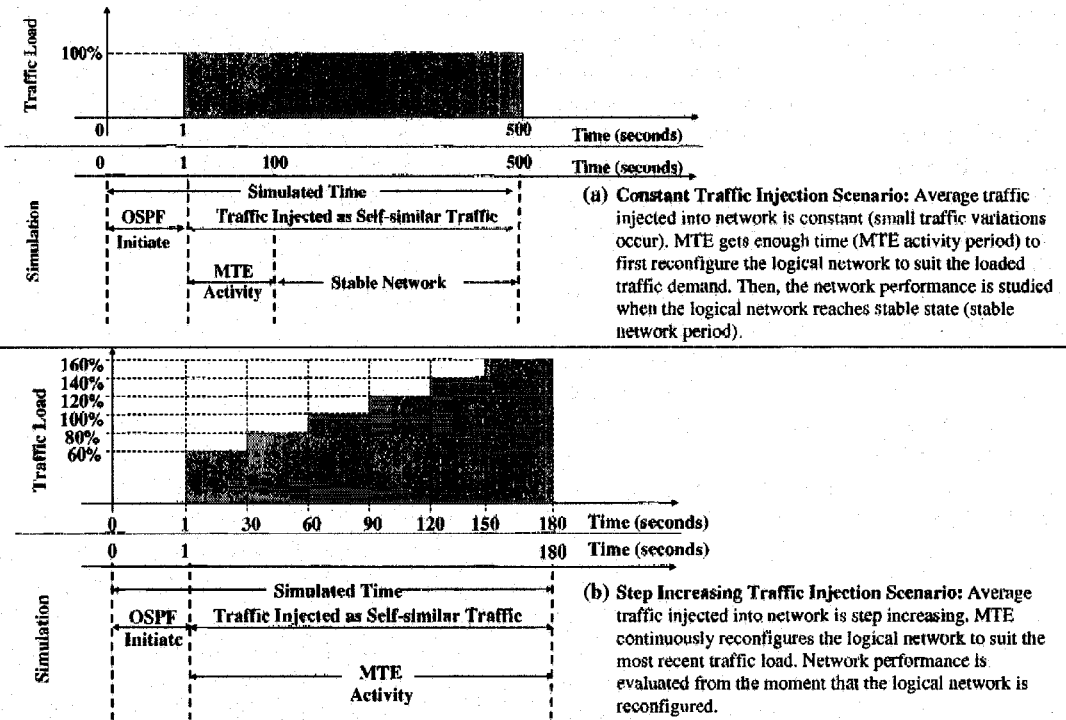


Fig. 11. Italian optical transport network topology.

The constant traffic injection scenario is illustrated in Fig. 12a. The traffic model used in this scenario is a random self-similar traffic model with a constant average demand. This scenario is used to evaluate the influence of the TOW length on the network performance in terms of packet loss, etc. This scenario runs for 500 s. Initially, the logical network topology is equal to the optical topology. Since OSPF is used as IP routing protocol, it takes some time to establish the routing tables. Considering the network size, in order to give enough time to initialize the IP routing method, the first second in every simulation instance is only used to initialize the network routing tables. This means no traffic is injected in the first second at all. After that, we start injecting traffic into the network; meanwhile the MTE model is switched on. MTE reconfigures the logical network topology to suit the actual traffic load. This period is denoted as MTE activity period. The network reaches a stable state after the MTE activity period. This means that from now on the

scenario (used in Section 4.3), and the other is a step increasing traffic injection scenario (used in Sections 4.4 and 4.5).



MTE Activity Period: Period that network is not yet in optimal configuration, traffic can still get lost
Stable Network Period: Duration of network performance comparison between MTE controlled network and static network
Applied Thresholds: $T_{low} = 30\%$, $T_{high} = 70\%$
100% Traffic Load: Traffic demand used to design an optimized static logical network
Network Performance Evaluation: Network performance in terms of packet loss ratio etc is evaluated every 100ms.

Fig. 12. Simulation scenarios.

logical network topology will not change much; only small adjustments to suit the most recent traffic load situation will be made. MTE monitors the traffic load and reconfigures the logical network topology if necessary (e.g., adds a new logical link when traffic congestion occurs). In case of a TOW of 3 s, the transient effect lasts for 25 s. Thus, in order to assure that the logical network reaches a stable state before we start evaluating the performance, the MTE activity period is set to 100 s.

The step increasing traffic injection scenario is illustrated in Fig. 12b. The traffic model used in this scenario is again a random self-similar traffic model, but this time with a step increasing average demand. This scenario is used to study the influence of the amount of available network resources on the MTE reaction and the network performance when the average traffic demand is increased in steps of 20%. For example, in Section 4.4, we will study the blocking ratio of the required optical connections when the traffic demand is increased. Just as in the constant traffic injection scenario, the first second in this scenario is used to initialize the routing tables in each node. Afterwards, the constant traffic pattern used in the constant traffic injection scenario is replaced by a step-by-step increasing traffic pattern, from 60% to 160% of the average demand that was used to design the optimized static logical network. Since this scenario aims to study the network performance when MTE reacts to the traffic increase, taking into account the optical network capacity (available wavelength channels), the average traffic demand increases with 20% of the constant average traffic demand every 30 s. MTE is applied to reconfigure the logical network to suit the most recent traffic load.

Finally, in both scenarios, logical links are torn down when they are detected to be underloaded in three consecutive TOWs.

4.2 Performance Comparison of an MTE-Enabled Network and A Static Network

In order to study the performance improvement realized by an MTE controlled network, the performance in terms of packet loss, etc. will be compared to that of a static network when we study the influence of the TOW length on the network performance (see Section 4.3) and evaluate the MTE performance in a network with

limited capacity (see Section 4.5). Since Section 4.4 aims to distinguish the influence of SPA and MPA on the QoS offered by an MTE-enabled network with limited capacity, this Section 4.4 does not consider the static case. The logical network topology used in the static case is optimized for the 100% traffic demand in order to achieve a minimal overall (IP and optical layer) network design cost. The applied grooming algorithm is described in [4]. This static logical network occupies 143 wavelength channels to make up 53 logical links. The topology stays fixed. In order to achieve a high wavelength channel filling (in that way reducing the cost) the static case does not use any high load threshold to limit the traffic load in a logical link. In other words, it is possible to fill a logical link up to 100% of its capacity.

From the simulation, we observed that the link-by-link grooming scheme is the dominant grooming scheme in the static case, which results in a longer buffer delay in the IP routers and an increase of the end-to-end delay. Moreover, since no capacity is set aside in the logical links (no T_{high}) to deal with the small traffic variations of the constant traffic demand scenario, the QoS is not as good as in the case of an MTE-enabled network (see Section 4.3). When a step increasing traffic demand is injected into the network (see Section 4.5), the QoS worsens fast in the overloaded situation since the static network was designed for a traffic load of 100%.

4.3 Performance Evaluation When Using Different TOW Lengths

The goal of this section is to study the influence of the TOW length on the network performance. Thus, the constant traffic injection scenario is applied and the network performance is evaluated after the network becomes stable (the period from 100 to 500 s in Fig. 12a). Every point illustrated in the charts in this section indicates the statistic value (average, minimum or maximum value) of an evaluated parameter over the stable period. The simulation results will be compared to the case of a static logical network. In each simulation instance, a different TOW length will be assigned. Finally, one important assumption to mention here is that the optical layer offers enough wavelength channels to satisfy the optical connection requests from

the MTE so that the network QoS will not be impacted by a failure in establishing the logical network.

Since the logical network is dynamic, the number of used IP links (a logical IP link may consist of one or more parallel wavelength channels) changes over time. Two kinds of links exist in the logical network: the stable links that will not be torn down after the network reaches the stable situation, and the dynamic links whose life-time depends on the traffic demand. The stable links, corresponding to the minimal set of links between the nodes in the logical network, are the backbone of the logical network. The dynamic links are the *ad hoc* links to relieve the congested links or in other words to recover from undesirable network situations. During the stable period, some links may be added to or removed from the logical network. This causes the number of logical links to vary over time. The minimum number of logical links indicates the situation with least connectivity in the logical network. Thus we consider the minimum number of logical links as the number of stable logical links. In Fig. 13a, the low end of the lines represents the minimum number of logical links, the high end of the lines represents the maximum number of logical links and the cross indicates the average number of logical links. Furthermore, the curve in Fig. 13b represents the number of logical network reconfigurations; and the curve in Fig. 13c illustrates the maximum number of total occupied wavelength channels (sum over all fibers).

In the case of a small TOW (0.1, 0.2, and 0.3 s), the logical network is reconfigured quite often. As

mentioned in Section 3, such small TOW lengths, i.e., smaller than the length of most traffic bursts, increase the system sensitivity. This means that each traffic burst can cause the logical network to be updated. Thus, sometimes, the number of links involved in the logical network is higher, which is represented by the higher maximum number of logical links. Nevertheless, the number of stable links is lower than in case of longer TOWs, which indicates that the number of logical links varies in a wide range. One may expect that the capacity efficiency will improve due to the sharing of physical resources. However Fig. 13c shows the opposite for the amount of required wavelength channels in the network. This is probably a result of the fact that the higher sensitivity requires the network to be able to cope with rather sporadic but high traffic peaks.

Increasing the TOW length (in the range from 0.5 to 3 s) improves the network stability, which is identified by the lower number of reconfigurations and a small variation range. This seems to be the most suitable TOW length range for the Italian IP-over-Optical network. Since large TOWs offer less sensitivity to the MTE system, the number of reconfigurations shown in Fig. 13b and the variance of the number of logical links (Fig. 13a) over time are reduced, which also implies an increasing number of stable links. Also, fewer wavelength channels are used in this case than in the case of a small TOW. This is explained by the lower sensitivity of the network, smoothing out any sporadic high peak in the traffic volume. On the other hand, we also observe that more wavelength channels are occupied when the TOW length is 1 s (refer to

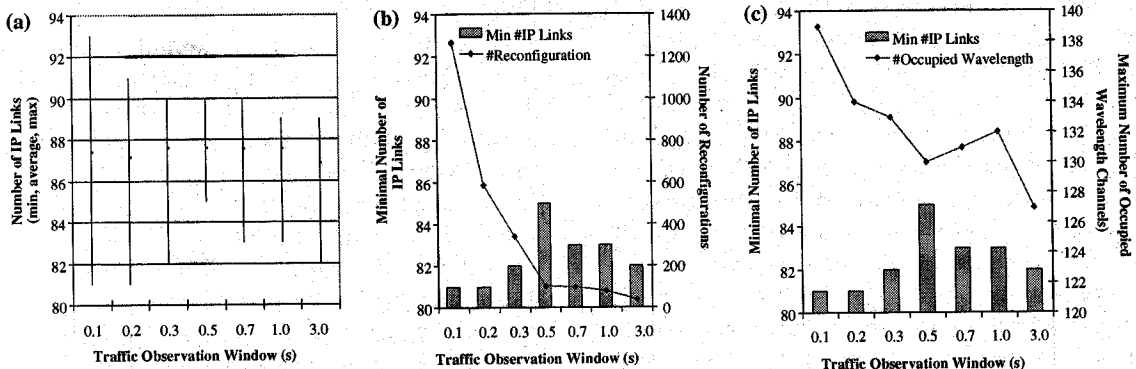


Fig. 13. Number of logical links and number of reconfigurations in the IP layer.

curve of maximum number of occupied wavelength channels in Fig. 13c). This is also a result of the decrease of the MTE system sensitivity. Taking into account that a logical network reconfiguration does not only imply setting up/tearing down a logical link, but also increasing/decreasing the capacity of an existing link, increasing the length of the TOW implies the reduction of the ability to adjust the capacity of existing links. Some links get a high capacity when a traffic burst is coming, but unfortunately, the capacity may not be decreased in time. This is the result of having an asymmetry between setting up and tearing down logical links. Tearing down a logical link requires three consecutive underloaded TOWs, which rarely happens in the case of a large TOW due to the insensitivity to the traffic variation. Therefore, the number of total occupied wavelength channels can be high when a large TOW is used. However, the number of occupied wavelength channels is dropped again when the 3 s TOW is used. In the case of the 3 s TOW, the high network stability (shown by few reconfigurations in Fig. 13b) is paid for by the reduction of sensitivity in detecting the traffic variation. This means MTE can neither add a new link to the logical network (lower average and maximum number of logical links shown in Fig. 13a) nor extend the capacity of an existing logical link. Thus, fewer wavelength channels are occupied in the case of the 3 s TOW. In the following study, we will see that such a large TOW also worsens the QoS (refer to Figs. 14 and 15).

To summarize the influence of the TOW length on the number of reconfigurations, on the number of stable IP links and on the total wavelength usage: When a small TOW is used, the logical network has few stable links, but requests more wavelength channels as it is frequently reconfigured. A suitable TOW (in the order of 0.5–1 s) improves the stability of the logical network, but uses fewer wavelength channels than with a smaller TOW. This means that each traffic flow has a high chance to get a direct logical link from source to destination. Meanwhile, few reconfigurations also imply that the MTE may not be able to extend or decrease the capacity of the logical network in time. This may result in keeping a lot of capacity in the logical network as well as more occupied wavelength channels. However, MTE may lose the ability to reconfigure the logical

network if the TOW is too big (as in the case of the 3 s TOW).

Since reconfiguring the logical network can change the traffic grooming situation, different TOW lengths may result in different traffic grooming states. We use the minimum number of logical links, the number of flows through a logical link, the average logical link utilization grade and the average buffer delay to evaluate the traffic grooming. Suppose the logical network has L logical links. Evaluating the number of flows through logical link i (F_i), the utilization grade of logical link i (U_i), and the buffer delay at the entrance of logical link i (BD_i) every 100 ms from 100 to 500 s, we can define:

Average number of flows through a logical link:

$$\bar{F} = \text{average} \left\{ \frac{1}{L} \sum_{i=0}^L F_{it} \mid 100 < t < 500 \right\} \quad (5)$$

Average logical link utilization grade :

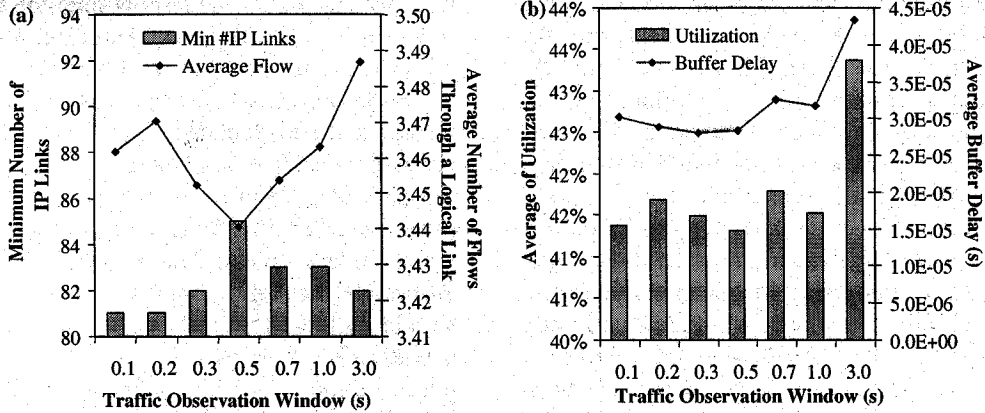
$$\bar{U} = \text{average} \left\{ \frac{1}{L} \sum_{i=0}^L U_{it} \mid 100 < t < 500 \right\} \quad (6)$$

Average buffer delay:

$$\bar{BD} = \text{average} \left\{ \frac{1}{L} \sum_{i=0}^L BD_{it} \mid 100 < t < 500 \right\} \quad (7)$$

In order to achieve minimum cost, the static case sets up less links in the logical network, which implies that the link-by-link grooming scheme is the dominant grooming scheme in the static case, and which results in a higher link utilization and higher number of flows per link comparing to the MTE-enabled network.

Since MTE dynamically reconfigures the logical network, the traffic grooming state is also dynamic. Using different TOW lengths may result in a different grooming state. In case of a short TOW (0.1–0.3 s), the network may contain many links when a traffic burst is coming, but most of these links are not stable. Thus, also the traffic grooming state is unstable: some traffic flows often swap between link-by-link grooming and end-to-end grooming, corresponding to a slight increase in the number of flows per link averaged over all links in the network (see

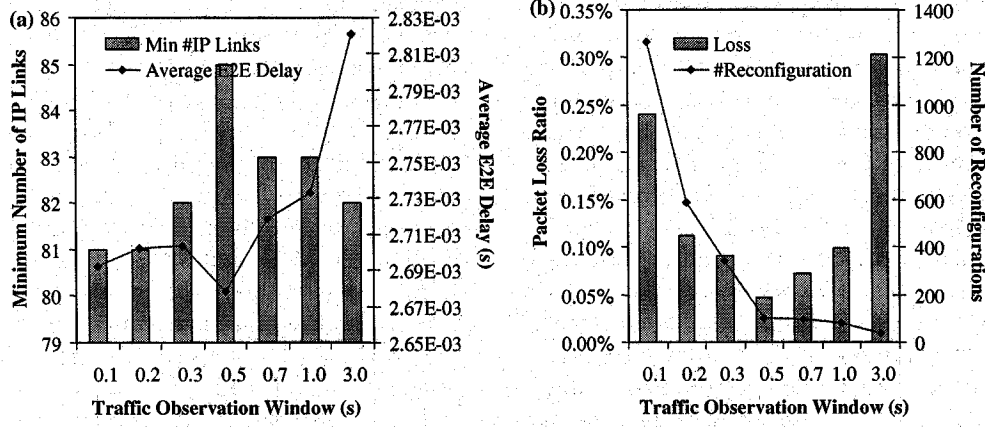


The average of the total number of links and the average number of flows through a logical link (a flow indicates a sequence of packets between a pair of IP routers.)

The average utilization of the logical links (in percentage of the link capacity) and the average length of the buffer delay in an IP router.

Static Case: Average number of connections per link: 7.04 Average IP link utilization: 76%

Fig. 14. Traffic grooming and IP link utilization.



Static Case: Average end-to-end delay: 2.86E-03(s) Packet Loss Ratio: 0.19%

Fig. 15. Offered QoS when using different TOW lengths.

Fig. 14a). By adding new links to the logical network, a traffic flow gets higher probability to be transported through a direct link from source to destination. Thus, due to the reduced traffic aggregation, MTE could avoid a high logical link utilization grade, which implies a lower packet loss ratio and a shorter buffer delay in the output ports. By enlarging the TOW a little bit (0.5–1 s), the average number of logical links and minimum number of logical links increase (see Fig. 13). This means that more logical links are stable and that the end-to-end grooming scheme becomes the

dominant grooming scheme in the network. As shown in Fig. 14b, each logical link contains the least number of flows as well as gets a low link utilization grade. However, if a quite large TOW is used, such as a 3 s TOW, the average number of logical links is small due to the reduction of sensitivity in reacting to the traffic variation. In this case, the importance of link-by-link grooming is increased. Thus, on average, each link contains more connections and has a higher link utilization (Fig. 14b), which may cause a higher buffer delay in the IP routers.

Fig. 15 illustrates two important QoS evaluation criteria: the average end-to-end delay and the packet loss ratio. Comparing to the MTE-enabled network, the static case does not have any capacity set aside (no T_{high}) to deal with the small traffic variations, thus packets have a higher probability to get lost during a traffic burst. On the other hand, since the link-by-link grooming scheme is the dominant traffic grooming scheme in the static case, the packets may stay longer in the buffer in the intermediate routers, which results in a longer end-to-end delay. As a result, the QoS offered by the static case is not as good as in the MTE-enabled network.

When a small TOW (0.1–0.5 s) is used to monitor the traffic demand, the offered QoS will degenerate due to the frequent reconfigurations. These network reconfigurations affect the average packet delay and the average packet loss ratio. Increasing the TOW length (from 0.5 to 1 s) not only increases the number of stable links, but also improves the network stability. This results in the increase of the average number of traffic flows passing through each logical link (see the curve of average number of flows through a logical link in Fig. 14a). When the TOW is further enlarged (3 s), due to the loss of sensitivity, MTE is no longer able to react timely to the traffic variations, although the logical network remains stable, the packet loss increases substantially. Furthermore, due to the low number of links in the logical network, link-by-link grooming becomes rather important. This results in the increase of the number of intermediate nodes along the route of an IP packet. Meanwhile, the increasing importance of link-by-link grooming also implies an increase of the utilization grade of each link and an increase of the buffer delay in the IP routers (refer to Fig. 14b). However, the increase of the buffer delay is insignificant.

Comparing the average IP buffer delay (shown in Fig. 14b) to the average end-to-end delay for an IP packet shown in Fig. 15a, the buffer delay is a very small part of the entire end-to-end delay. This is because MTE mainly results in end-to-end grooming, which reduces the competition on the network resources between the traffic flows and reduces the number of intermediate nodes. Consequently, an IP packet gets a higher probability to get a direct link from source to destination and

may not (or for a rather short time) be buffered in intermediate IP routers. Therefore, the average buffer delay in the IP routers becomes small.

To summarize, when decreasing the TOW length, more direct end-to-end logical links are established in the network (i.e., end-to-end grooming grows in significance). However, the network becomes more unstable: the frequent reconfigurations will affect the end-to-end delay and packet loss ratio. Although the higher logical network dynamism should allow physical capacity sharing (i.e., wavelength channels on the optical fibers), this phenomenon is compensated/dominated by the fact that the network should be able to accommodate very high but sporadic peaks in the traffic volume.

4.4 Influence of The Underlying Layer Routing Methods

As we know, shortest path routing may cause heavy wavelength occupation in some part of the underlying physical network. Thus, optical connections requested by the MTE may sometimes get a high blocking probability. In order to reduce the blocking ratio when establishing an optical connection, we proposed two routing methods in Section 2.3: MPA and SPA. The goal of this Section is to investigate the influence of these two routing methods on the network performance and the MTE reactions when the network is under different traffic loads.

The results presented here are for two case studies on a realistic Italian national OTN under the same traffic load situation. The first one assumes that each fiber has 5 wavelength channels; the second assumes that each fiber has 15 wavelength channels. The capacity of each wavelength channel is 2.5 Gbps. The step increasing traffic injection scenario, as shown in Fig. 12b, is used in this section to obtain different traffic load levels for the simulated network illustrated in Fig. 11. The performance of the MTE system is evaluated in terms of the blocking of the extra logical links needed to accommodate the additional traffic, the average number of intermediate nodes used per additional optical connection established when the traffic load is increased, and the packet loss ratio during each simulation instance. The two possible underlying

layer routing methods (SPA or MPA) will be compared to investigate their impact on the MTE actions.

Fig. 16 clearly shows the impact of a limited number of wavelength channels. As the offered traffic volume grows from 60% to 160%, more and more optical connections, which are required by the extra logical links needed to accommodate the additional traffic, will be blocked, indicating an increasing capacity shortage. However, in the case of 15 wavelength channels per fibers, no optical connections will be blocked, since sufficient capacity is available to accommodate the largest traffic volume. Fig. 16 also shows that MPA slightly outperforms SPA. This was expected from the example in Fig. 8.

Fig. 17 confirms the results presented in Fig. 16. First of all, on average longer optical connections are seen with SPA, confirming the slightly higher blocking probability. Although a larger traffic volume typically results in more direct and thus

longer optical connections (i.e., the slight increase in Fig. 17 for 15 wavelength channels per fiber), this effect seems insignificant compared to that caused by the capacity shortage (i.e., steep slope for 5 wavelength channels per fiber).

Finally, Fig. 18 illustrates the perceived packet loss ratio, a measure to estimate the QoS. In accordance with the vast amount of blocked optical connections for a traffic volume larger than 140% in the case of 5 wavelength channels per fiber (see Fig. 16), the logical IP network perceives a continuous excessive (>1%) loss of traffic. In other circumstances, blocking is less frequent or even very sporadic, leading to some short spikes in the charts of Fig. 18.

In Fig. 18, we also observed that about 2% more packets get lost while Fig. 16 shows almost 30% more blocking of the additional logical links when the traffic demand is increased from 140% to 160% in the case of 5 wavelength channels per fiber. The reason is that the set up of a new logical link is

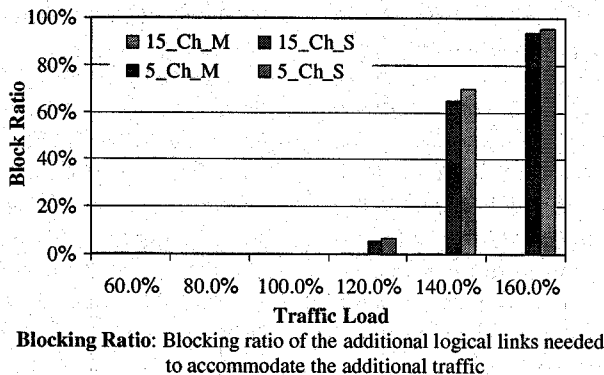


Fig. 16. Blocking ratio of the optical connections. The labels are formatted as X_Ch_Y: X indicates the number of wavelength channels per fiber, and Y indicates the routing method: M stands for MPA, S for SPA.

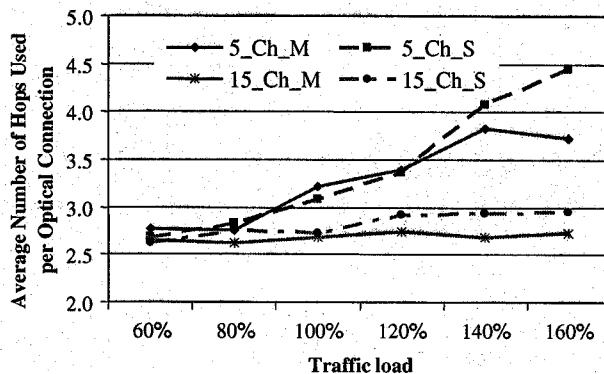


Fig. 17. Average length in hops of the additional optical connections needed to accommodate the additional traffic.

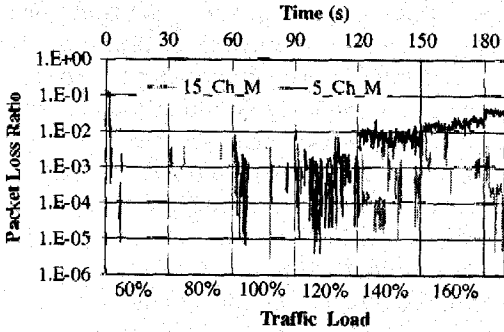


Fig. 18. Packet loss ratio over 100 ms.

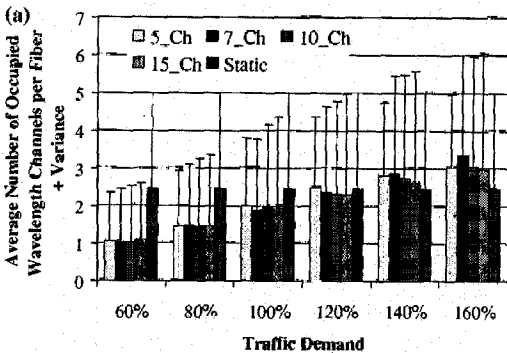
already triggered when the load on an existing logical link exceeds T_{high} . This implies that the logical link has still some capacity left to accommodate this and possibly a (slightly) higher traffic load.

To summarize the results discussed in this section, we can say that the ability of STE may be restricted in networks with insufficient capacity since the required optical connections get a high blocking ratio due to the lack of capacity, as shown in Fig. 16. We have seen that MPA achieves a lower blocking probability in establishing optical connections in the underlying physical layer. Compared to SPA, the improvement is small. On the other hand, MTE also causes an increase of the optical connection path length in hops (see Fig. 17) when the network capacity becomes too small, which may lengthen the end-to-end delay for packets. Finally, MTE may lose its advantage in improving the QoS, (for example, the packet loss ratio as shown in Fig. 18) if the network capacity becomes too small.

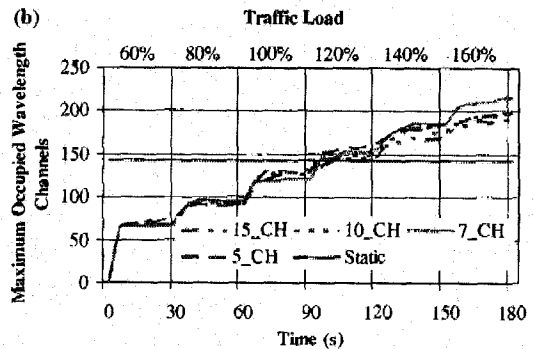
4.5 Performance Evaluation for A Network with Limited Wavelength Channels

In the previous section, we have introduced the advantage of MPA. This section investigates the performance of an MTE-enabled network under different traffic load situations. Therefore, the step increasing traffic injection scenario is used in each simulation. The TOW is set to 0.5 s. MPA is used to route the optical connection in the underlying layer. Since MPA may split a required optical connection into several physical paths, we assume in the following discussing that the delay and hop-length of an optical connection is the delay and length of its longest physical path.

Fig. 19a presents the average number of used wavelength channels per fiber and its variance during a certain traffic load level (each traffic level stays for 30 seconds). The bars indicate the average number of wavelength channels used per optical fiber in both the dynamic and the static optimized case. When the traffic volume is increased from 60% to 160%, the average number of occupied wavelength channels in a fiber triples in all dynamic cases. In the network with fibers containing more channels, like the case of 15 channels per fiber, the variance of the occupied wavelength channels in the fibers doubles. For the network with fewer channels per fiber, like the case of 5 channels per fiber, the variance only increases with 40%. A high variance means that the wavelength channel occupation in the underlying OTN is not symmetric; some fibers may stay at a higher utilization grade, others not. Furthermore, since the static logical network was designed for the traffic



Average wavelength channel occupancy per fiber under different traffic loads (each demand stays 30 seconds)



Maximum of the total occupied wavelength channels over one second

Fig. 19. Wavelength channel occupancy.

demand of 100% and aimed at high link filling grades, it has little extra capacity in the logical links to avoid QoS degradation during traffic bursts. In contrast, MTE reconfigures the logical network dynamically to avoid QoS degeneration during traffic bursts, meanwhile using the network resources more efficiently. This is why we observed in Fig. 19b a lower amount of wavelength channels requested by MTE for those traffic loads under 100%.

In addition, Fig. 19b presents the increment of the total wavelength channel usage in the optical network. We evaluated the total wavelength channels usage every 100 ms. Putting all evaluated values in a figure would make it unreadable. Therefore, we only illustrated the maximum of total number of occupied wavelength channels over 1 s in Fig. 19b. For the same reason, Figs. 21, 22 and 23b also illustrate the statistic results (average values) over 1 s. When MPA is used by MTE, the total number of occupied wavelength channels in the network with fibers containing few channels (refer to the trace of 7 channels per fiber) could be higher than that in the network with fibers containing more channels (refer to the trace of 15 channels). However, if the wavelength channels are too few to satisfy the requests from the logical layer, most of the requested optical connections will be blocked in the optical layer due to a lack of wavelength channels, e.g., the case of 5 channels per fiber. This makes it impossible to extend the logical network capacity, resulting in less logical links in the logical network in Fig. 20a.

Therefore, its total wavelength channel usage is as low as the case of 15 channels per fiber. However, as Figs. 21 and 22 present, the QoS is unacceptable.

Fig. 20 illustrates the traffic grooming result at different traffic load situations. As shown in Fig. 20b, since the static case is optimized for 100% traffic load, the traffic grooming scheme is fixed. This could result in the longer buffer delay presented in Fig. 23. On the contrary, MTE timely reconfigures the logical network to cope with the incremental traffic load by increasing the number of links in the logical network. Therefore, the traffic flows get a higher possibility to use an end-to-end direct link in the logical network. This results in the reduction of the number of aggregated traffic flows in a logical link. Comparing the number of traffic flows through an IP link (Fig. 20b) between an MTE-enabled network and the static network, this is much lower in the MTE controlled case. This proves again that the main traffic grooming scheme used by MTE is end-to-end grooming. The importance of this grooming scheme increases with the traffic load. In other words, MTE prefers link-by-link grooming under lower traffic load in order to reduce the number of occupied wavelength channels; nevertheless, increasing traffic demand will result in an increased importance of the end-to-end grooming scheme.

After evaluating the network resource usage and the traffic grooming state for a network with limited capacity, we are going to study the network performance (in terms of packet loss and

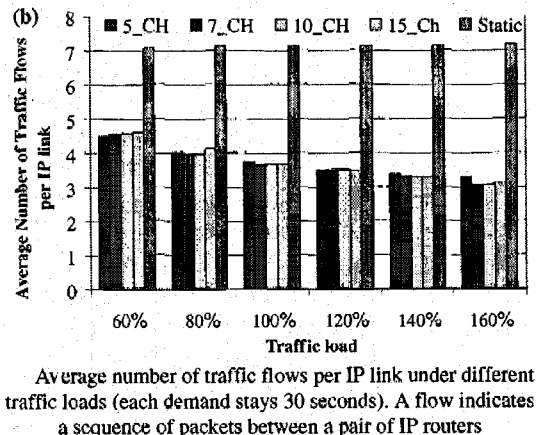
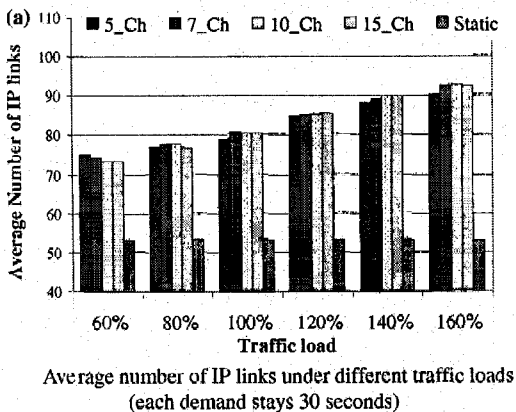


Fig. 20. Traffic grooming in networks with limited capacity.

delay) in an MTE-enabled network with limited capacity.

Firstly, let us consider the packet loss ratio of networks with different amounts of available capacity and under different traffic loads. Although MTE uses the wavelength channels more efficiently compared to the static case, we can still ask the question whether it is possible to keep a packet loss in the networks with fewer wavelength channels per fiber that is as low as in the networks with more wavelength channels per fiber. Fig. 21 illustrates the curve of the packet loss ratio in the static case and the dynamically controlled cases for different numbers of wavelength channels per fiber.

Although the packets get a lower loss ratio in the static case when the traffic load is below 100%, more packets get lost under higher traffic loads. This is due to the fixed logical network capacity in the static case. In contrast, using MTE the packet loss in the networks with more than 5 wavelength channels per fiber can also be reduced in case the network load is heavily loaded. In the case of 5 channels per fiber, the packet loss is reduced, but remains unacceptable when the traffic volume gets higher. Since the capacity in the network is too small, the network has to deal with wavelength channel shortage of all over the network. Consequently, the optical network cannot satisfy the requests from the MTE. However the case of 7 channels per fiber results in a much better packet loss ratio than the case of 5 channels per fiber. On one hand, this is because the additional wavelength channels per fiber increase the probability of success in establishing an optical connection. On the other hand, MPA tries to satisfy the request by using the fibers outside the congested area, which results in the higher average wavelength usage in each fiber and the small variance in the number of occupied wavelength channels over the network (Fig. 19). Thus, in the case of 7 channels per fiber, the packet loss ratio remains acceptable and approaches the cases of 10 or 15 channels per fiber.

Secondly, we are going to evaluate the average IP packet delay both in terms of the end-to-end delay and the delay due to the buffer in the IP routers.

Unlike the static case, where a fixed number of wavelength channels is occupied over time, MTE

uses, as long as the network is lightly loaded, a logical network with low capacity. This is represented by the few occupied wavelength channels in Fig. 19. The logical network contains few IP links (see Fig. 20) meaning that the link-by-link grooming approach becomes the dominant grooming scheme (note the large average number of flows per IP link in the 60% loaded networks in Fig. 20), which may result in a longer buffer delay in the IP routers. The IP router buffer delay increases the end-to-end delay in all MTE controlled cases. With the increase of the traffic demand, the static network becomes congested, which is presented by the high packet loss ratio in Fig. 21 and the long IP router buffer delay in Fig. 23b. Thus, its end-to-end delay and delay jitter increase. The end-to-end delay in the static case becomes even longer than that in all dynamical cases (refer to Fig. 22).

To review the dynamic cases, the end-to-end packet delay in the networks with fewer wavelength channels, i.e., the case of 5 or 7 channels per fiber, is increasing fast as the network gets more heavily loaded. There are two reasons for this increasing end-to-end delay. The first one is the increase of the hop-length of an optical connection. This can be observed in Fig. 23a. Because of the lack of wavelength channels in the fibers, the system may have to set up the additional optical connections, meant to satisfy the connection requirement from the higher layer in order to alleviate the network congestion, through a path which can be far away from the congested area. This results in a longer propagation delay. Secondly, the required connection may still be blocked in the optical layer: Fig. 20a for instance implies that the logical network cannot be extended due to the lack of wavelength channels in the case of 5 channels per fiber. Thus, the traffic congestion in the existing logical link cannot be solved, which results in a higher buffer delay in the intermediate routers (refer to Fig. 23). Vice versa, we observed that the end-to-end delay in the dynamic cases with a large amount of capacity is shortened when the network is heavily loaded. This is the advantage of having more end-to-end grooming and a large amount of capacity in the underlying optical network. Firstly, more end-to-end grooming offers more direct links to the traffic flows, which implies that little demand is transferred back to the logical network in

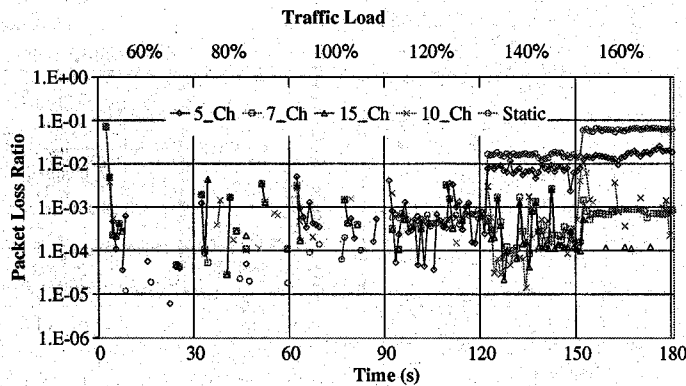


Fig. 21. Average packet loss ratio over 1 s in networks with a limited number of wavelength channels per fiber.

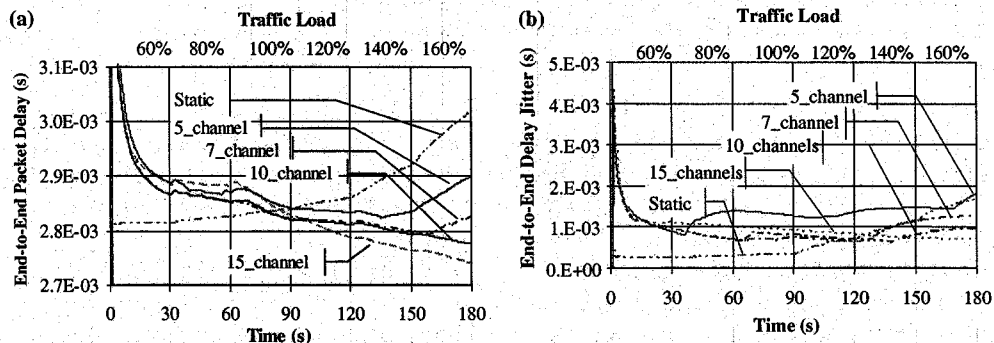


Fig. 22. Average packet end-to-end delay and delay jitter over 1 s in networks with limited wavelength channels.

intermediate nodes along the traffic flow's path, as well as a slight reduction of the buffer delay in the IP routers as shown in Fig. 23b. Secondly, the underlying optical network is able to satisfy the requested wavelength channels, so that the traffic gets high probability to pass through a direct link from source to destination (presented as more IP links and less flows through a certain IP link in Fig. 20). Consequently, the time needed to hand over a packet from the optical layer to the IP layer in the intermediate nodes for routing purposes is avoided, which also implies a reduction of the end-to-end delay.

Since the IP layer topology is dynamically reconfigured by the MTE, the end-to-end path may be changed frequently, so that the dynamic cases also encounter higher delay jitter in Fig. 22b. To compare the trace of the delay jitter between the different dynamic cases, the less wavelength channels the optical fiber contains, the larger the packet end-to-end delay jitter is. This is reasonable

because the requested optical connection gets higher probability to be rejected in the optical network with fewer wavelength channels than in the network with more wavelength channels. This results in the failure in reconfiguring the logical network when congestion occurs. Thus, each logical link has to aggregate more traffic flows (refer to Fig. 20); and network capacity starts fluctuating and the variation of end-to-end delay is enlarged.

5 Conclusion

In this paper, we introduced a reactive MTE strategy that automatically and dynamically reconfigures the upper layer logical network by properly updating the optical connections in the underlying physical layer. This reactive MTE strategy applies a TOW to measure the traffic load in the logical links. When an overloaded situation is detected in a TOW, MTE is triggered to set up a

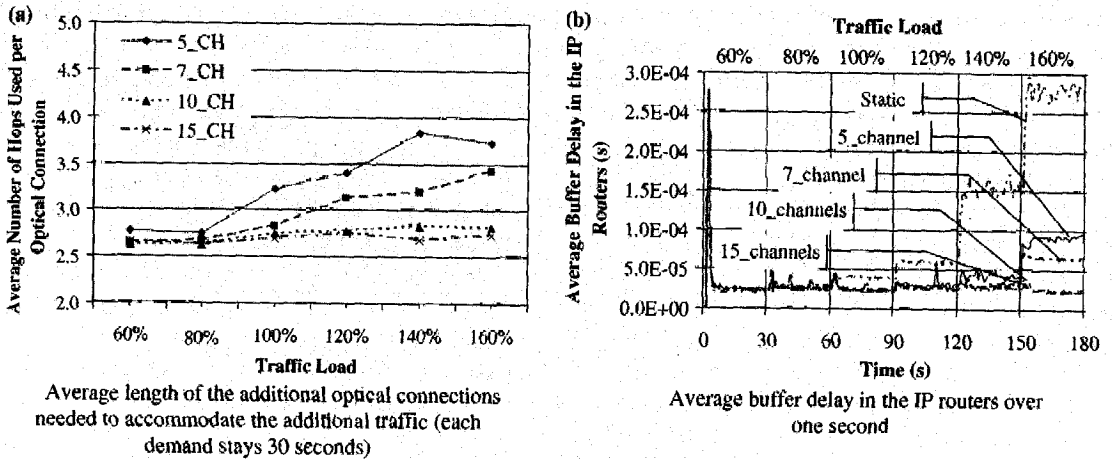


Fig. 23. The length (in hops) of the additional optical connections and the average buffer delay in the IP routers per second (hop-length here is the hop-length of the longest path of an optical connection).

new logical link to cope with the traffic bursts. However, tearing down an existing logical link requires three consecutive underloaded TOWs. Different TOW lengths can lead to different actions to deal with the same traffic pattern.

After introducing the technical details used by the MTE strategy, the results of a simulation on a practical case study were presented to demonstrate the influence of the TOW length on the network performance. For a typical traffic pattern, a small TOW causes the MTE to be oversensitive to the traffic variations and results in unstable logical links as well as in unnecessary logical link tear-down actions, which can destroy the network stability and result in an insufficient QoS. A suitable TOW can duly extend the logical network capacity by setting up a stable logical link, which results in a stable network and a good QoS over time. Finally, a large TOW does not react accurately enough to the traffic variations; although the network stays stable, packets get higher loss probability.

We also observed that using different TOW lengths could lead to different traffic grooming states. If the TOW is small, the logical network is not stable and also the traffic grooming situation is unstable. A suitable TOW can lead to a preferred stable state in which the low rate flows are groomed link-by-link and the high rate flows are groomed end-to-end. Enlarging the TOW will increase the importance of link-by-link grooming. Thus, for the case of a big TOW, the link-by-link

grooming scheme becomes the dominant grooming scheme, which may lead to longer buffer delays in the IP routers and also lengthen the end-to-end packet delay.

If the capacity in the underlying physical network is limited, an enhanced routing algorithm able to reduce failures in establishing the optical connections due to a lack of available wavelength channels. In the simulation results, the MPA presents a slight advantage over the SPA when establishing an optical connection in the underlying physical network with lower capacity. However, if the underlying physical network capacity is too low, the MTE may not be able to keep a good QoS when traffic bursts come, since the required optical connections are blocked in the underlying layer due to a lack of capacity.

Acknowledgments

The authors would like to thank the IWT and ITEA for supporting this work through the TBONES project, the IWT for supporting this work through the GBOU "Optical Networking and Node Architectures" project and the European Commission for supporting this work through the IST project NOBEL. B. Puype and D. Colle would like to thank the IWT for supporting their work through respectively a Ph.D. and a post-doctoral scholarship.

References

- [1] E. Modiano, P. J. Lin, Traffic grooming in WDM networks, *IEEE Communication Magazine*, vol. 39, no. 7, (July 2001), pp. 122–133.
- [2] R. Dutta, G. Rouskas, Traffic grooming in WDM networks: past and future, *IEEE Network*, vol. 16, no. 6, (Nov/Dec 2002), pp. 46–56.
- [3] K. Zhu, B. Mukherjee, Traffic Grooming in WDM mesh network, *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 1, (Jan. 2002), pp. 110–121.
- [4] S. De Maesschalck, M. Pickavet, D. Colle, P. Demeester, Multi-layer traffic grooming in networks consisting of an IP/MPLS layer on top of a meshed optical network layer, *Proc. of Globecom'03* (San Francisco, USA, Dec. 2003).
- [5] ITU-T Rec. G.8080, Architecture for the automatically switched optical network (ASON), (Nov 2001).
- [6] OIF working groups of Architecture, OAM & P, PLL and Signaling, User Network Interface (UNI) 1.0 Signaling Specification, OIF2000.125.5 (June 2001).
- [7] B. Jamoussi et al., Constraint-Based LSP Setup using LDP, Internet Network Working Group, <ftp://ftp.rfc-editor.org/in-notes/rfc3212.txt> (Jan. 2002).
- [8] P. Trimintzios et al., Engineering the multi-service Internet: MPLS and IP-based techniques, *Proc. of ICT'01*, vol. 3 (Romania, Bucharest, June 2001), pp. 129–134.
- [9] X. Xiao, L. M. Li, Internet QoS, a big picture, *IEEE Network Magazine*, vol. 13, no. 2, (March/April 1999), pp. 8–18.
- [10] S. Nelakuditi, Z. L. Zhang, A localized adaptive proportioning approach to QoS routing, *IEEE Communications Magazine*, vol. 40, no. 6, (Nov 2002), pp. 66–71.
- [11] M. Kodialam, T. V. Lakshman, Restorable dynamic quality of service routing, *IEEE Communications Magazine*, vol. 40, no. 6, (June 2002), pp. 72–81.
- [12] X. Masip-Bruin, S. Sanchez-Lopez, J. Sole-Pareta, J. Domingo-Pascaual, Reducing routing inaccuracy effects by using the BYPASS based routing mechanism (University Politecnica de Catalunya, 2002), <http://www.ac.upc.es/reerca/reports/DAC/2002/index.en.html> #UPC-DAC-2002-30.
- [13] K. Sato, GMPLS-Based photonic multilayer router (Hikari Router) architecture: An overview of traffic engineering and signaling technology, *IEEE Communication Magazine*, vol. 40, no. 3, (March 2002), pp. 96–101.
- [14] J. Moy et al., OSPF Version 2, Network Working Group (April 1998).
- [15] D. Awduche et al., RSVP-TE: Extensions to RSVP for LSP Tunnels, RFC 3209, Internet Draft Network Working Group (Dec. 2001).
- [16] Y. Ye, C. Assi, S. Dixit, M. A. Ali, A simple dynamic integrated provisioning protection scheme in IP-over-WDM network, *IEEE Communications Magazine*, vol. 39, no. 11, (Nov. 2001), pp. 174–184.
- [17] B. Puype et al., Multi-layer traffic engineering in data-centric optical networks, *Proc. of ONDM'03* (Budapest, Hungary, Feb. 2003), pp. 221–226.
- [18] Q. Yan et al., Applying intelligent optical network functionality in the deployment of the next-generation internet, *Proc. of Annual Symposium of the IEEE/LEOS Benelux* (Amsterdam, Netherlands, Dec. 2002), pp. 59–62.
- [19] J. Ash, M. Girish, E. Gray, B. Jamoussi, G. Wright, Applicability statement for CR-LDP, RFC 3213, Internet Draft Network Group (Jan. 2002).
- [20] The Network Simulator: ns-2, <http://www.isi.edu/nsnam/ns>.
- [21] S. Robert, J. Y. Le Boudec, New model for pseudo self-similar traffic, *Performance Evaluation*, vol. 30, no. 1–2, (July 1997), pp. 57–68.

Bart Puype received his M.Sc. degree in electro-technical engineering in 2002 from Ghent University, Belgium. Since then he has been working as a research assistant, preparing his Ph.D., at the Department of Information Technology of Ghent University. In 2004 he received an IWT scholarship. His main interests are in the field of communication networks, focusing specifically on the design and evaluation of multi-layer IP-over-Optical networks. He is currently involved in the IW1/ITEA TBONES and IST NOBEL projects (bart.puype@intec.UGent.be).



Didier Colle received a M.Sc. degree in electrotechnical engineering (option: communications) from the Ghent University in 1997. Since then, he has been working at the same university as researcher in the department of Information Technology (INTEC). He is part of the research group INTEC Broadband Communication Networks (IBCN) headed by prof. Piet Demeester. His research leads to a Ph.D. degree in February 2002. From January 2003 on, he was granted a postdoctoral scholarship from the "Instituut voor de aanmoediging van Innovatie door Wetenschap en Technologie in Vlaanderen (IWT-Vlaanderen)". His research deals with design and planning of communication networks. This work is focusing on OTNs, to support the next-generation Internet. Up till now, he has actively been involved in several IST projects (LION, OPTIMIST, DAVID, STOLAS, NOBEL and LASAGNE), in the COST-action 266 and 291 and in the ITEA/IWT TBONES project. His work has been published in more than 70 scientific publications in international conferences and journals.



Ilse Lievens received an M.Sc. degree in electrical engineering, focusing on telecommunications, in 1994 from the Ghent University. She then joined the Department of Information Technology at the Ghent University, where she obtained a Ph.D. degree in 2000 in the Broadband Communication Networks Group (IBCN research group). Her Ph.D. thesis "Use of Distributed Rerouting in Meshed ATM Networks" looked at the design of rerouting algorithms for survivability in meshed ATM networks, focusing on distributed, autonomously working techniques. She is currently working as a post-doctoral assistant in the IBCN research group. Her research interests involve broadband communication networks, focusing on the design, reliability and survivability in IP and optical networks. She is and has been involved in several European projects (ACTS, IST, and ITEA), and national inter-university projects. She is author and co-author of several publications in conference proceedings and journals (ECOC, IEEE JSAC, IEEE Commun. Magazine, PNC,...).



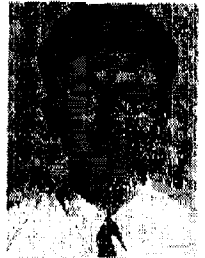
Mario Picquet received an M.Sc. and Ph.D. degree in electrical engineering, specialized in telecommunications, from Ghent University in 1996 and 1999, respectively). Since 2000, he is professor at Ghent University where he is teaching telecommunication networks and algorithm design. His current research interests are related to broadband communication networks (WDM, IP, (G-)MPLS, OPS, OBS) and include design, long-term planning and routing of core and access networks. Special attention goes to Operations Research techniques that can be applied for routing and network design. In this context, he is currently involved a.o. in the European IST projects "Switching Technologies for Optically Labeled Signals" (STOLAS), "All-optical Label Swapping employing Optical Logic Gates" (LASAGNE) and the Network of Excellence on optical networks ePHOTON/One. He has published several international publications, both in journals (e.g., IEEE JSAC, IEEE Comm. Mag., Eur. Trans. on Telecommunications, Photonic Network Communications,...) and in proceedings of conferences.



Piet Demeester is professor at the Ghent University, where he is involved in research on communication networks. His current interests are related to broadband communication networks (IP, G-MPLS, optical packet and burst switching, access and residential, active, mobile, CDN, grid) and include network planning, network and service management, telecom software, internetworking, network protocols for QoS support, etc. He published over 300 journal or conference papers in this field.



Qiang Yan received a B.S. degree in telecommunication engineering in 1997 from Shanghai University, China. Since 1998 he has been working as a research assistant at the Department of Information Technology of the Ghent University, preparing a Ph.D. His main research interests include the design, planning and evaluation of IP networks and IONs. Up to now, he was involved in the IST project LION and COST action 266.



Sophie De Maesschalek received a M.Sc. degree in Electro-technical Engineering (option Communication Technology) in 1998 from the Ghent University, Belgium. Since 1998 she has been working as a research assistant at the Department of Information Technology of the Ghent University. In 2003 she obtained a Ph.D. degree from the Ghent University, where she is currently working as a post-doctoral researcher. Her main research interests include the design, planning and evaluation of multi-layer IP-over-OTNs and broadband access networks. She was involved in the ACTS project SONATA, in the IST project LION and in COST action 266, and she is currently working on the European projects BREAD and MUSE.

