A Traffic Engineering System for Multilayer Networks Based on the GMPLS Paradigm

Paola Iovanna, Roberto Sabella, and Marina Settembre, Ericsson Lab Italy

Abstract

This article discusses a novel approach for realizing traffic engineering in the framework of new-generation multilayer networks based on the GMPLS paradigm. In particular, the proposed traffic engineering system is able to dynamically react to traffic changes while at the same time fulfilling QoS requirements for different classes of service. The proposed solution consists of a hybrid routing approach, based on both offline and online methods, and a novel bandwidth management system that handles priority, preemption mechanisms, and traffic rerouting in order to concurrently accommodate the largest amount of traffic and fulfill QoS requirements. The bandwidth resources of the network are effectively exploited by means of "elastic" utilization of the bandwidth. The main building blocks and operations of the system are reported, and the major advantages are discussed.

It is widely recognized that traffic will be more and more dominated by Internet-based services, with respect to traditional voice traffic [1], thanks to increased adoption of high-speed access technology and migration of more and more services toward the Internet Protocol (IP). Voice traffic is still growing, but at a slower rate. As a result, two main factors are of critical importance in the development of new-generation networks (NGNs): the sheer quantity of traffic is growing rapidly, and the type of traffic is changing.

As a result the telecommunications world is evolving strongly toward challenging scenarios: the convergence of the *telecom* and *datacom* worlds into the *infocom* era is becoming a reality. New infrastructures have to be compliant with such an infocom network scenario. In practice this means that network infrastructure has to be multiservice, that is, able to support several types of traffic with different requirements in terms of quality of service (QoS) [2]. Since IP traffic will be the dominant portion, network infrastructures must take into account its characteristics. Two main attributes typify Internet traffic:

- Its self-similar nature
- Asymmetry of the data flows

As a whole, Internet traffic is not easily predictable and stable as is traditional voice traffic. Consequently, a basic requirement arises for new-generation infrastructure: flexibility and ability to react to traffic demand changes with time.

Another key issue relates to the fact that even though Internet traffic is becoming dominant, it does not generate revenue as do valuable voice services. This, practically, means that if the network were upgraded by adding bandwidth and expanding infrastructure in proportion to the amount of data traffic increase, the revenues would be smaller than the cost. Thus, in order to be profitable, Internet service providers (ISPs) and network carriers must both reduce costs by means of an effective use of network resources and increase revenues by offering multiservice and QoS capabilities.

Moreover, the migration of all services over IP, including the real-time ones, requires guaranteeing QoS for a subset of services that should be comparable to those provided by telecom-based networks nowadays.

As a result, several requirements come out for NGNs: provide fast provisioning, handle traffic fluctuations and growth, handle the QoS to honor service level agreements (SLAs) for different types of traffic in terms of bandwidth, delay, packet loss, or any other quality requirements, and offer multiservice capabilities.

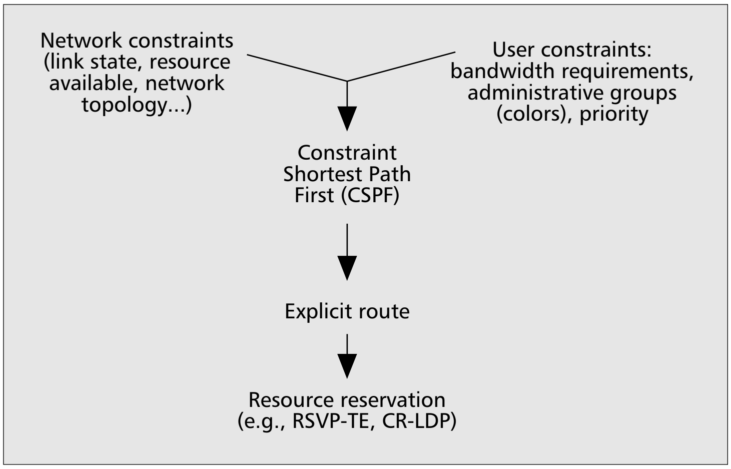
The challenging task for established network operators is how to migrate their voice network toward the new-generation infrastructure, while minimizing the costs of the transition and taking early advantage of the benefits offered by next-generation networks.

Multiprotocol label switching (MPLS) addresses these issues by means of traffic engineering (TE) mechanisms that allow the advantages of flexibility and performance in conjugating layers 3 and 2, respectively [3]. The challenge for NGNs consists of extending such flexibility and efficiency to other layers of the network, such as synchronous digital hierarchy/synchronous optical network (SDH/SONET) and wavelength-division multiplexing (WDM) in order to consider even non-packet-based forwarding planes.

Thanks to the MPLS extension by means of generalized MPLS (GMPLS), the key ingredients to perform efficient TE for different technologies are available [4]. However, a feasible solution that is able to use such ingredients is still not consolidated. Actually, TE should provide the network with the possibility to dynamically control traffic data flows, to optimize the availability of resources, to choose routes for traffic flows while taking into account traffic loads and network state, and to move traffic flows toward less congested paths. All these functions should be performed handling different network layers and technologies.

This article describes a pragmatic network solution that

0890-8044/03/$17.00 © 2003 IEEE



■ Figure 1. *The principle of constraint-based routing.*

addresses the above-mentioned issues by exploiting the generalized version of an MPLS network model in a multilayer scenario, and is able to support QoS and bandwidth on demand services. Such a solution has been developed in our laboratory through a testbed that makes use of extended versions of IP signaling and routing protocols. The goal of this article is to propose an innovative solution, describe its building blocks and modes of operation, and discuss its characteristics. The technical details and performance of the main building blocks are beyond the scope of the present article.

The article is organized as follows. In the next section the network scenario and technical background are described. We explain the proposed solution, specifying the main building blocks that allow the realization of TE in the multilayer network and the TE system operations in response to different events. The characteristics of the proposed TE system are discussed, and finally some conclusions are derived.

The Network Scenario and Technical Background

Traffic engineering is the process to control traffic flows in the network in order to optimize resource use and network performance [5, 6]. Practically, this means choosing routes taking into account traffic load, network state, and user requirements such as QoS and bandwidth, and moving traffic from more congested paths to less congested ones. In order to achieve TE in an Internet network context, the Internet Engineering Task Force (IETF) introduced MPLS [7], constraint-based routing [8], and enhanced link state interior gateway protocols (IGPs) [9, 10] as key ingredients. Actually, it is widely known that an MPLS control plane together with proper constraint-based routing (CBR) solutions provide the means for achieving TE, thus allowing the provisioning of new services based on the bandwidth-on-demand concept, such as flexible virtual private networks (VPNs).

MPLS

MPLS architecture is a standardized structure able to support advanced TE solutions and QoS functionalities. It is based on the separation between data plane and control plane, reusing and extending existing IP protocols for signaling and routing functions, while reintroducing a connection-oriented model in an Internet-based context [11]. The MPLS scheme is based on the encapsulation of IP packets into labeled packets that are forwarded in an MPLS domain along a virtual connection called a label switched path (LSP). MPLS routers are called label switched routers (LSRs), and the LSRs at the ingress and egress of an MPLS domain are edge LSRs (E-LSRs). Each LSP can be set up at the ingress LSR by means of ordered control before packet forwarding. This LSP can be forced to follow a route that is calculated a priori thanks to the explicit routing function. Moreover, MPLS allows the possibility to reserve network resources on a specific path by means of suitable signaling protocols, such as Resource Reservation Protocol with TE (RSVP-TE) or Constraint-Based Routing with Label Distribution Protocol (CR-LDP) [11]. Thus, the LSP represents a virtual connection in the MPLS network like virtual circuits and virtual paths in the asynchronous transfer mode (ATM) world.

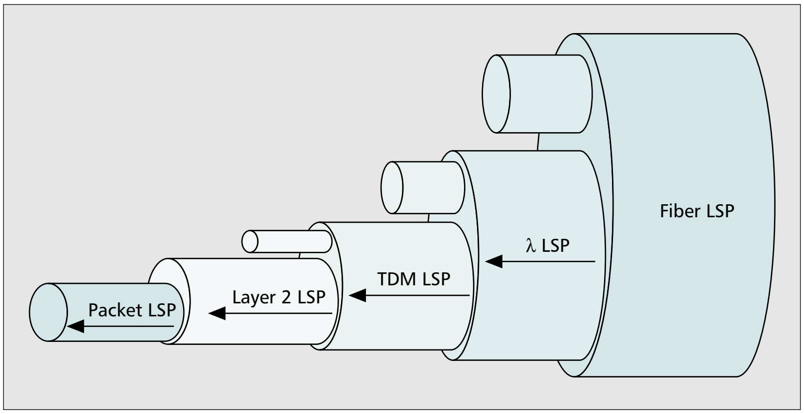
In particular, each LSP can be set up, torn down, rerouted if needed, and modified by means of the variation of some of its attributes, including the bandwidth [6]. In fact, the bandwidth of an LSP can be modified dynamically, just for the desired increment [12], according to a specific request at the ingress LSR preserving all the other attributes.

Furthermore, preemption mechanisms on LSPs can also be used in order to favor higher-priority data flows at the expense of lower-priority ones, while avoiding congestion in the network. Another important feature of MPLS relates to the possibility of stacking labels, providing the means to introduce different hierarchical levels instead of the two provided by ATM [11]. This feature favors VPN services support and, as is clarified later, allows extending MPLS control to other technologies.

Constraint-Based Routing

The combination of the explicit routing function, resource reservation mechanisms, and CBR in the MPLS network represents the key to an efficient TE strategy [8]. In particular, the criteria utilized to choose routes in a network and possibly to reroute traffic flows toward alternative paths are crucial for applying TE strategies. Such criteria necessarily take into account more parameters than simply network topology. A simple sketch of CBR operations is shown in Fig. 1. In fact, when calculating the route for a requested path (LSP in the case of MPLS-based networks), CBR has to take into consideration both network and user constraints. The former regards the link state and resource availability besides network topology, while the latter relates to bandwidth requirements, administrative groups, priority, and so on. When an explicit route has been computed, the resource reservation procedure is started by means of signaling protocols such as RSVP. In this way, CBR may find longer but less congested paths instead of heavily loaded shortest paths. Thus, network traffic is distributed more uniformly and congestions are prevented.

Two main approaches can be considered for calculating routes: offline and online. Basically, the offline approach refers to a predetermined route computation, usually accomplished by an external network optimization tool (e.g., an external server), while the online approach refers to an "on demand" route computation, automatically achieved by means of signaling protocols or an external tool. The offline approach is adequate for achieving global path optimization on the basis of a traffic matrix that represents the foreseen connection requests for any pair of network nodes. Such a traffic matrix is usually derived by a statistical expectation of traffic demands. Logically, this method is quite appropriate when traffic demand is reasonably stable; traffic changes are not so important as to require a redesign of the routes for the different data flows. This is the case of traditional voice traffic that is quite predictable and reasonably stable, and thus the traffic matrix is quite consistent. Unfortunately, Internet traffic is



■ Figure 2. *LSP hierarchy in GMPLS.*

neither predicable nor stable. Therefore, a pure offline approach could be inadequate since it could lead on one hand to wasted network resources (the transmission pipes are not filled) or, on the other hand, to congestion because the amount of traffic is increased and the assigned resources are not enough. To promptly react to Internet traffic changes an online approach could be more satisfactory. In particular, the online routing method consists in evaluating the route on demand, when needed (i.e., when there is a new request or a change of a previous request). Thus, it is suitable to perform a single LSP accommodation at a time. The main problem in those cases is to preserve the stability. In fact, instability can occur when the time necessary to route a new data flow is on the order of the period of time in which the requests are originated. Clearly, the online approach is also inadequate to perform global path accommodation. Moreover, online routing may lead to higher resource consumption and is not scalable. As a result, a hybrid approach could be the best solution in order to exploit the advantages of both methods.

From the above considerations, it emerges that CBR in real networks is a crucial and complex issue.

In this article we propose a pragmatic TE system that utilizes an innovative hybrid routing approach. More specifically, the TE system invokes an offline procedure to achieve global optimization of path calculation, according to an expected traffic matrix, while invoking an online routing procedure to dynamically accommodate, sequentially, actual traffic requests, thus allowing prompt reaction to traffic changes. As is described in more detail later, the original contribution of the proposed hybrid routing solution consists of the integration of the two routing functions. Such functions can be realized in different ways, without affecting the applicability of the solution. Clearly, the ways the two routing functions are achieved have an impact on system performance, for example, in terms of accommodated traffic amount.

The GMPLS Paradigm for New Generation Networks

To extend the features of the MPLS technique, the generalized version of it (GMPLS) presents a gradual and future-proof approach toward NGNs [4, 13, 14]. In practice, the GMPLS control plane can manage heterogeneous network elements (e.g., IP/MPLS routers, SDH/SONET elements, ATM switches, and even optical elements) using a suitably extended version of the well-known IP protocol suite. This makes possible the realization of a single control plane able to handle a whole multilayer network. In particular, GMPLS extends the MPLS concepts even to non-packet-switched technology by means of the LSP forwarding hierarchy [15]. This is shown in Fig. 2 [14]. The GMPLS forwarding hierarchy is based on the multiplexing capabilities of the node interfaces. At the top of such a hierarchy (external LSP in the figure) are nodes that have fiber-switch-capable interfaces (i.e., fiber cross-connects); at the second stage (λLSP in the figure) are nodes with wavelength switching capabilities (i.e., optical cross-connects, OXCs); at the third stage (TDM LSP) are nodes with TDM switching capabilities (e.g., SDH cross-connects); at the fourth stage (layer 2 LSP) are nodes with layer 2 switching capabilities (e.g., real MPLS routers or ATM switches); and at the last stage (packet LSP) are nodes with packet switching capabilities (e.g., IP routers). Any stage can be associated with a network domain that can be nested into another one. The outer domain represents the packet LSP domain. The layer 2 LSP domain is nested inside the packet one and so on up to the inner domain representing the fiber LSP one. It is to be highlighted that each LSP should be generated and terminated on homogeneous devices (i.e., belong to the same network domain). On the other hand, a packet-switch-capable LSP can be nested and tunneled into an already existing higher-order LSP.

GMPLS can support different network scenarios, where heterogeneous layers can cooperate in several ways for the convenience of manufacturers and operators. Without losing generality, we consider a two-layer network as a reference scenario, consisting of an IP/MPLS layer, whose network elements are basically LSRs, and a WDM transport layer, whose nodes are OXCs, as depicted in Fig. 3. Specifically, just packet LSPs and λLSPs are considered. The latter represent end-to-end optical connections or lightpaths.

Interworking between the IP/MPLS and optical layers is another key issue. Particularly, the packet-based structure of the IP/MPLS layer and the circuit-based construction of the optical layer have to be harmonized. This means that any lightpath bundles several LSPs, characterized by different bandwidth attributes. The bandwidth attribute of each LSP belonging to the IP/MPLS layer varies over a continuous range, while the lightpath bandwidth is fixed to the wavelength channel bit rate.

Different deployment scenarios can be envisaged for optical networks based on GMPLS concepts, with overlay and peer as the extremes [4]. Each of them defines a different level of interworking between the IP/MPLS and optical layers. The overlay model is based on a client-server approach. In this context, the optical layer acts as a server of the IP/MPLS layer. The control planes are separated in this case and communicate with each other by means of a standard user–network interface (UNI) [16]. In this case the IP/MPLS network asks for a connection, and the optical network manages its resources in order to set it up according to the SLA. In the peer model a single control plane manages the whole network. In this way all the nodes, both the IP/MPLS and optical ones, act as peers sharing the same complete topological view. This allows a network operator to have a single domain composed of different network elements, providing greater flexibility. The price for this is the amount of information that has to be handled by any network element. The deployment scenario has an impact on routing strategy. In particular, two main strategies can be adopted in a GMPLS-based network: single-layer and multilayer. In a single-layer approach, the LSPs are aggregated by edge LSRs into lightpaths. At this point the connection requests are expressed in terms of the number of wavelengths requested by each pair of optical nodes. The optical layer is then responsible for finding the routes for the optical LSPs and assigning the wavelengths (i.e., solving the routing and wavelength assignment, RWA, problem) [17, 18]. In a multilayer approach the aggregation and routing are jointly performed, allowing an LSP to be routed on a concatenation of lightpaths in a single routing instance and leading to efficient use of network resources [19, 20]. It is logical that awareness of the status of all network elements and the possibility to manage the whole set of network resources allow more efficient routing functions to be performed.

Regarding QoS handling, GMPLS can reserve bandwidth for individual LSPs at any hierarchical level.

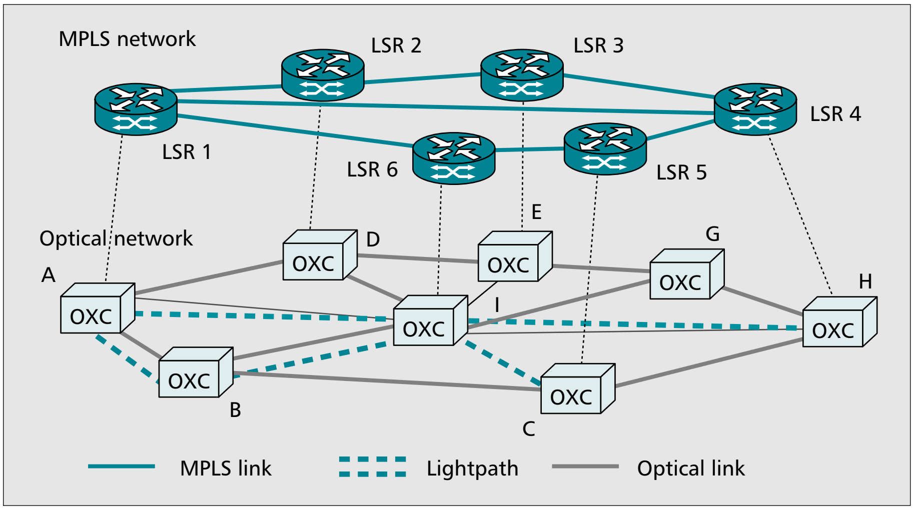
The potential of a GMPLS control plane in terms of advanced TE capabilities, provided by cooperative interworking among layers, is remarkable, but the feasibility of a simple and effective TE solution is still challenging. In effect, different technological and architectural aspects have an impact on the practical implementation of TE strategies, in terms of:

Complexity of the CBR function: The realization of such a function, taking into consideration simultaneously all network and user constraints in a network made of many and heterogeneous network elements, is very complex. Thus, a simple and practical approach is needed for routing.

QoS handling: Managing different QoS requirements for several classes of services in the network is another complex task. Specifically, this deals with ways to achieve traffic segregation, routing according to different priority levels, and preemption.

Signaling: In order to be efficient, the CBR has to know the updated link state of the whole network, and possibly the map of all the LSPs. This means a huge information flood through the network. Therefore, it is necessary to find a reasonable trade-off between routing efficiency and amount of information to be flooded throughout the network.

Prompt reaction to traffic changes: In principle, the network should be able to react to traffic changes promptly. This requires the possibility of realizing dynamic routing of data flows according to such requests. This could also lead to stringent technological requirements for the nodes at all levels. Moreover, online routing leads to nonoptimal routes compared to global routing performed offline. Thus, it is necessary to find a reasonable combination of dynamic routing facilities and static routing.



■ Figure 3. *The multilayer reference network scenario.*

Enhanced Signaling

The development of GMPLS requires a suitable extension of MPLS signaling and routing protocols in order to manage heterogeneous technology [14].

This means that the routing protocols, such as Open Shortest Path First with TE (OSPF-TE), have to perform flooding of detailed and updated topology information and attributes for each link at different network layers, and signaling protocols such as RSVP-TE have to handle the generalized label concept to support the establishment of LSPs at any hierarchical level [21]

As a result, extended routing and signaling protocols have to cope with a huge and heterogeneous amount of information in respect to a pure MPLS-based network, leading to scalability issues. For instance, the overall number of links in an optical network can be several orders of magnitude bigger than in an MPLS network. To address such an issue the concept of link bundling has been introduced [4]. In fact, similar optical parallel links can be aggregated to form a bundle for routing purposes. On the other hand, the signaling of each individual component of the bundle requires a new protocol, introduced specifically for link management in the optical networks, called Link Management Protocol (LMP) [4, 22]. Specifically, LMP is responsible for:

- Establishing and maintaining control channel connectivity
- Verifying the link physical connectivity
- Rapidly identifying link, fiber, and channel failure within the optical domain

However, the efficiency of CBR depends not only on the amount of disseminated information on network topology and resource availability, but also on the frequency of information updating. The more detailed and up-to-date the information collected in the link state database, the better the routing decision is likely to be. Dynamic link state routing suffers this problem, especially due to the fast changeability of the constraints to be considered.

These issues could be addressed by inheriting mechanisms already used in the MPLS network, such as threshold methods that avoid excessive flooding or methods based on a timer setting an upper bound on flooding frequency [23]. However, in a GMPLS scenario, these mechanisms could be insufficient to make a completely dynamic link state routing approach feasible. This is a further reason to use a hybrid routing approach, as mentioned earlier. In this way the dependence of route computation on flooding information is relaxed.

A Traffic Engineering System for New-Generation Optical Networks

The main requirements for a TE system of an NGN can be summarized as follows:

- Optimize the use of network resources (e.g., link bandwidth and node throughput) by means of "elastic" use of the bandwidth resource.
- Actualize the bandwidth-on-demand concept.
- Support different classes of service (CoSs), including real-time traffic (e.g., the CoS foreseen in the differentiated services, DiffServ, scenario defined by IETF) and guarantee the required QoS.

The basic idea of the proposed TE system lies in a hybrid routing approach, based on both offline and online methods, and bandwidth management systems that allow QoS requirements to be fulfilled.

Specifically, since the offline procedure is not subjected to strict computational time requirements and does not need information dynamically disseminated by routing protocols, it is convenient to adopt a multilayer approach that allows optimization of network resources, as explained earlier. Due to the fact that traffic changes are not easily predictable and can appreciably vary the traffic distribution itself, it is necessary to use online procedures that allow new requests to be sequentially accommodated on demand. This function is realized by the dynamic routing function, which takes as input individual connection requests, and attempts to route them in such a way as to prevent congestion. Besides the routing functions, the TE system makes use of bandwidth modifying mechanisms foreseen by the MPLS model in order to provide the bandwidth to a given connection just for the time it is actually requested, aiming to improve resource utilization (i.e., flexibility) in the network. In fact, bandwidth modifying mechanisms allow the bandwidth attribute of any LSP to be varied according to specific requests. If modify operations are permitted (i.e., no congestion is foreseen along the considered path), TE allows the bandwidth attribute to be modified. When bandwidth is reduced, the portion of bandwidth that is released is put at the disposal of the network in order to accommodate new requests. If bandwidth is increased, the TE provides more bandwidth to that connection while maintaining the old route for that path. If the modify operation is not allowed, TE can decide to either reject that request or reroute that connection on another path. The dynamic routing function is therefore used to either accommodate online new traffic requests that can be originated by unpredicted demands, or reroute some portion of the traffic in order to prevent congestion in those cases when bandwidth modification cannot be done, preserving the old routes. In this way, the bandwidth resource throughout the network is effectively used according to traffic demands, and all the connections are associated with an "elastic" bandwidth attribute that can increase or decrease according to the specific request. However, if QoS has to be assured, it is necessary to introduce some mechanisms to handle possible congestion and solve contention among different requests. For this reason it makes sense to introduce some priority mechanisms in order to assign resources to higher-priority LSPs at the expense of lower-priority ones when needed. For instance, lower-priority LSPs could be preempted if they consume network resources needed by higher-priority LSPs.

The considered TE system makes use of priority mechanisms to distinguish among different classes of services and handle network resources in order to manage such priorities. Furthermore, the proposed TE system is able to assure the bandwidth to all those connections that cannot tolerate any degradation of QoS parameters. From now on we will refer to such connections as *premium* paths. In order to do that, this system assigns the route, during the path provisioning phases, to the entire set of premium LSPs, providing them the maximum bandwidth attribute those connections could require during their life, and makes use of a specific component that is able to make those routes available in any traffic condition. This component, the bandwidth engineering (BE) module, operates in an attempt to optimize the use of network resources and prevent congestion by rerouting a selected portion of LSPs that occupy the bandwidth required by higher-priority LSPs and, if it does not succeed in finding alternative routes for such LSPs, preempting them.

It is useful for practical purposes to distinguish between two main groups of LSPs. The first type of LSP relates to the premium traffic, and can be referred to as highest priority (HP) LSPs. The second relates to all other types of lower-priority LSPs and can be referred to as LP LSPs. This second group could be further classified into several classes (e.g., LP1, LP2, etc.) according to level of priority. While the HP LSPs are guaranteed at any time and in any traffic condition, whatever their bandwidth attribute up to the maximum allowable value agreed, in megabytes, by the SLA, all the LP LSPs are not guaranteed and compete among themselves according to the different levels of priority. For instance, an LP1 LSP can preempt an LP2 or LP3 LSP, and so on.

In this way both HP and LP traffic is served on demand, but HP traffic routes are precalculated during the path provisioning phase and fixed, while LP traffic routes can be dynamically changed according to different load conditions and priority policy. Specifically, LP LSPs can be rerouted or even preempted to prevent congestion, according to their level of priority.

In all, the proposed TE system consists of an efficient integration of the different building blocks performing the path provisioning function (offline routing), dynamic routing (online routing), and the BE function that is able to actualize the elastic bandwidth concept. Such an integrated solution provides flexible and dynamic utilization of network resources in order to face a consistent variation of traffic distribution due to the unpredictability of Internet traffic and traffic demands varying with time, and concurrently to accommodate the largest amount of traffic while guaranteeing the desired bandwidth attribute for premium connections at any time, whatever the traffic demand. Obviously, the performance of the TE system depends on the specific implementation of the different building blocks. To better explain how the system operates, it is useful to discuss the main building blocks that constitute the TE system and describe the main events handled by the TE system.

Building Blocks

The TE system utilizes three main building blocks for its operations:

- A path provisioning module (PR)
- A dynamic routing module
- A BE module

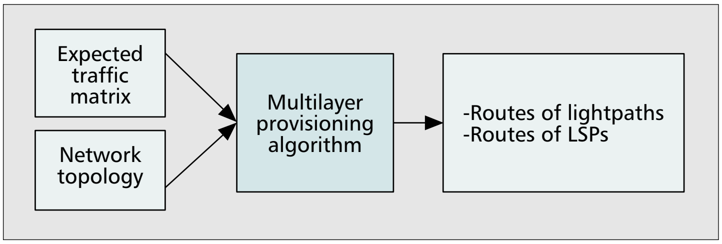
Before explaining how TE works in normal operations and how it reacts to relevant events, it is worth explaining how the key building blocks perform.

The building blocks constituting the TE systems are listed here.

Path Provisioning Module (PR) — The path provisioning module action is illustrated in Fig. 4. It calculates offline the routes for all foreseen connections, according to a traffic matrix that describes the traffic relationships between each network node pair, on the basis of the physical topology of the network and information about network resources (e.g., presence of wavelength conversion inside the OXCs, link capacity). The traffic matrix, which accounts for different types of traffic, is evaluated by the operator on the basis of either the agreements stipulated with clients or the estimation made through statistical evaluation. In a two-layer network architecture the global path provisioning problem can be schematized in two steps:

- Design a logical topology of the optical layer, that is, set the lightpaths (i.e., the λLSPs) and their physical routes
- Routing the LSPs at the IP/MPLS layer onto the logical topology

Typically two subproblems are separately performed: first, the LSPs are suitably groomed according to a given objective function (e.g., the cost of electronic and optical multiplexing devices); then lightpath provisioning is achieved on the basis of a traffic matrix expressed in terms of number of wavelengths. These approaches can be regarded as single-layer [17,18]. The proposed system operates in a multilayer fashion, by simultaneously solving grooming and routing of LSPs, and RWA of



■ Figure 4. *Sketch of the provisioning module.*

optical paths (i.e., lightpaths). Thus, close interworking between the MPLS and the optical layers is realized, as mentioned earlier. The path provisioning algorithm has an objective function that must fulfill two criteria:

- Minimize the number of lightpaths in order to optimize data flow aggregation
- Minimize the number of lightpaths an LSP spans during its travel throughout the network to reduce the number of times it is electronically processed inside the nodes (LSRs) [19]

In order to allow the DR module to easily react to traffic changes, it may be opportune to introduce a suitable overprovisioning at the optical level. Specifically, by limiting the bandwidth of each wavelength constituting the lightpaths, more lightpaths are set up in the provisioning phase. As a result, the DR module can operate on a logical topology provided by the PR module, which is enforced more just where the traffic is expected to be [24].

Dynamic Routing Module — The DR module evaluates the route for a single LSP request at a time, expressed in terms of source and destination nodes and bandwidth requirements. Route computation is performed considering the actual link state status of both the MPLS and optical layers, which is learned by flooding of routing protocols such as extended OSPF. Basically, the DR algorithm finds a route aimed at better utilizing network resources by using less congested paths instead of shortest, but heavily loaded paths. In order to find the route, the DR algorithm has to fulfill at the same time two criteria:

- Finding a route so that the traffic is evenly distributed on the MPLS layer
- Bundling the LSP onto the lightpaths to increase the probability of finding available wavelengths for subsequent connections demanding even large bandwidth

This means that the DR algorithm favors the choice of less congested routes that contain less loaded links at the MPLS layer, and chooses more occupied wavelengths at the optical layer in order to efficiently aggregate LSPs into lightpaths. Specifically, the DR accomplishes this by means of a proper weight system that takes into account not only the number of hops, but also the capacity available in any link and on individual wavelengths [20]. Even the online routing procedure applies the same considerations made for the PR module about interworking between the MPLS and optical layers. Thus, the grooming and routing functions are simultaneously accomplished. A simple and fast realization of a DR module can operate using just the logical topology of the optical networks provided by the PR module. It means that the DR cannot set up new lightpaths. In this case, suitable overprovisioning, mentioned in the above PR module description, facilitates the task of DR. In fact, a limited increase of network resources could lead to a significant increase of performance of DR even in critical loading conditions [24]. The opportunity of setting up new lightpaths dynamically may be worth investigating in a future work.

Bandwidth Engineering Module — The TE system is based on elastic use of bandwidth: the bandwidth can be temporarily released by higher-priority LSPs and put at disposal of all the lower-priority LSPs. This can be done provided that the bandwidth is immediately given back to high-priority traffic as soon as needed. Therefore, a function is needed to handle preemption of lower-priority LSPs or, even better, to move lower-priority traffic onto less congested routes. In fact, when a higher-priority LSP requires more bandwidth and at least one link on its path is congested, the BE module is invoked to make the required bandwidth available. The most rudimentary BE module can be represented by a preemption module that tears down all the LSPs whose priority level is lower than that of the LSP to be accommodated. An advanced version of a BE module consists of a system that uses a priority policy to select the LSPs to be removed and tries to reroute them on alternative paths, and eventually tears down those paths it does not succeed in rerouting [25]. In fact, the BE module contains:

- An algorithm to properly select the LSPs to be removed in an attempt to minimize the amount of traffic to be torn down
- A dynamic routing algorithm that can be the DR module itself

The BE module is invoked anytime there is a need to prevent congestion on a certain route.

Other key elements of the TE system are the databases where all the information required is recorded. In principle, three basic information components are needed:

Routes database (RDB): The RDB contains all the routes calculated offline by the provisioning module. For each route the source-destination nodes pair, classes of service, client identification, and bandwidth are also specified. The bandwidth value read in the RDB refers to:

- The MB value in case of HP flows as set by the SLA
- The value considered in the traffic matrix, which can represent either the average or minimum bandwidth according to the network operator policy

TE database (TED): It contains the status of each link and its attributes (e.g., available bandwidth, reserved bandwidth) and is continuously updated by means of information flooding achieved through routing protocols (e.g., OSPF-TE).

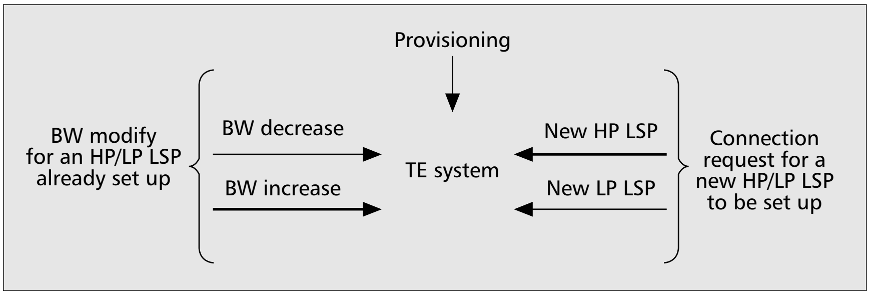
Dynamic LSP database (DLD): It reports detailed information on the status and attributes of each current LSP in terms of source-destination pair, route, classes of service, and bandwidth.

TE System Operations

In order to better understand how the TE system works, several possible events have to be taken into consideration. In the following we describe the different events and how the TE system reacts to those events, as illustrated in Fig. 5.

The considered events are listed below.

Global Path Provisioning Request — In traditional telecom infrastructures, global optimization of LSP routes as performed in the provisioning phase happens quite rarely. In NGNs this may occur anytime there is a significant change in traffic distribution. For instance, the introduction of a new ISP in the network area providing a different pricing policy or new services may lead to a significant variation of traffic distribution, or the network operator (or the carrier) establishing new contracts with old or new customers may require redesign of the traffic flows in the network. Such external events trigger the provisioning module. In this condition the PR module operates, finding an optimal solution for all the routes relating to all the CoSs. The PR module provides the routes to individual LSPs, possibly aggregating them in bigger data flows such as wavelength channels (or lightpaths).


Provisioning
BW modify for an HP/LP LSP already set up
BW decrease
BW increase
TE system
New HP LSP
New LP LSP
Connection request for a new HP/LP LSP to be set up

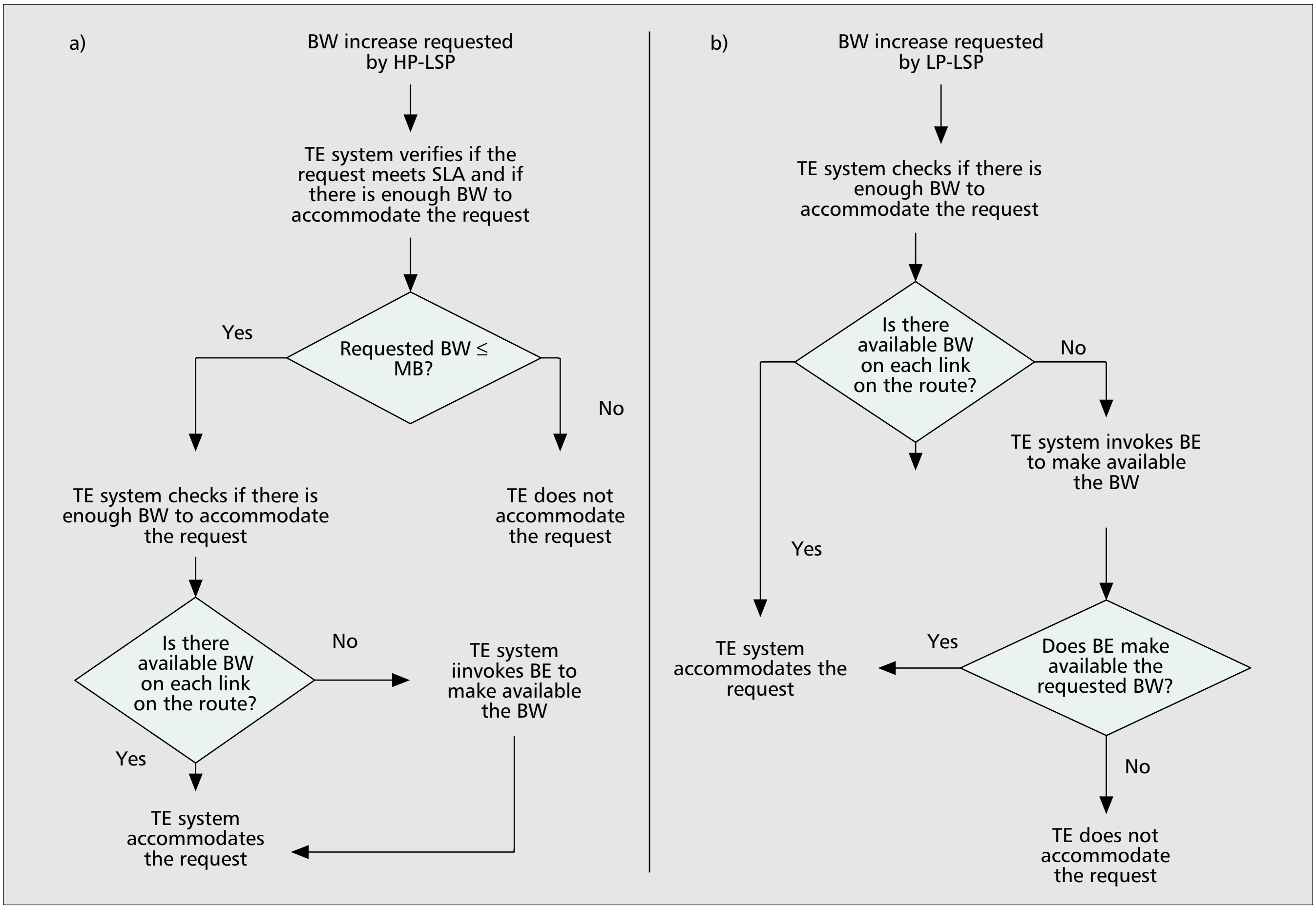
■ Figure 5. *Events handled by the TE system.*

Bandwidth Decrease Request — Any LSP can request to decrease its bandwidth attribute in a certain period. If the network operator can manage this situation, the advantage for the client is that he/she can pay less, because he/she consumes less bandwidth, while the advantage for the network operator is that it can use the network resource to serve other traffic requests. In this case the TE achieves the bandwidth modification according to known MPLS mechanisms, and updates the relevant information in the databases, thus making available the released bandwidth to accommodate new requests.

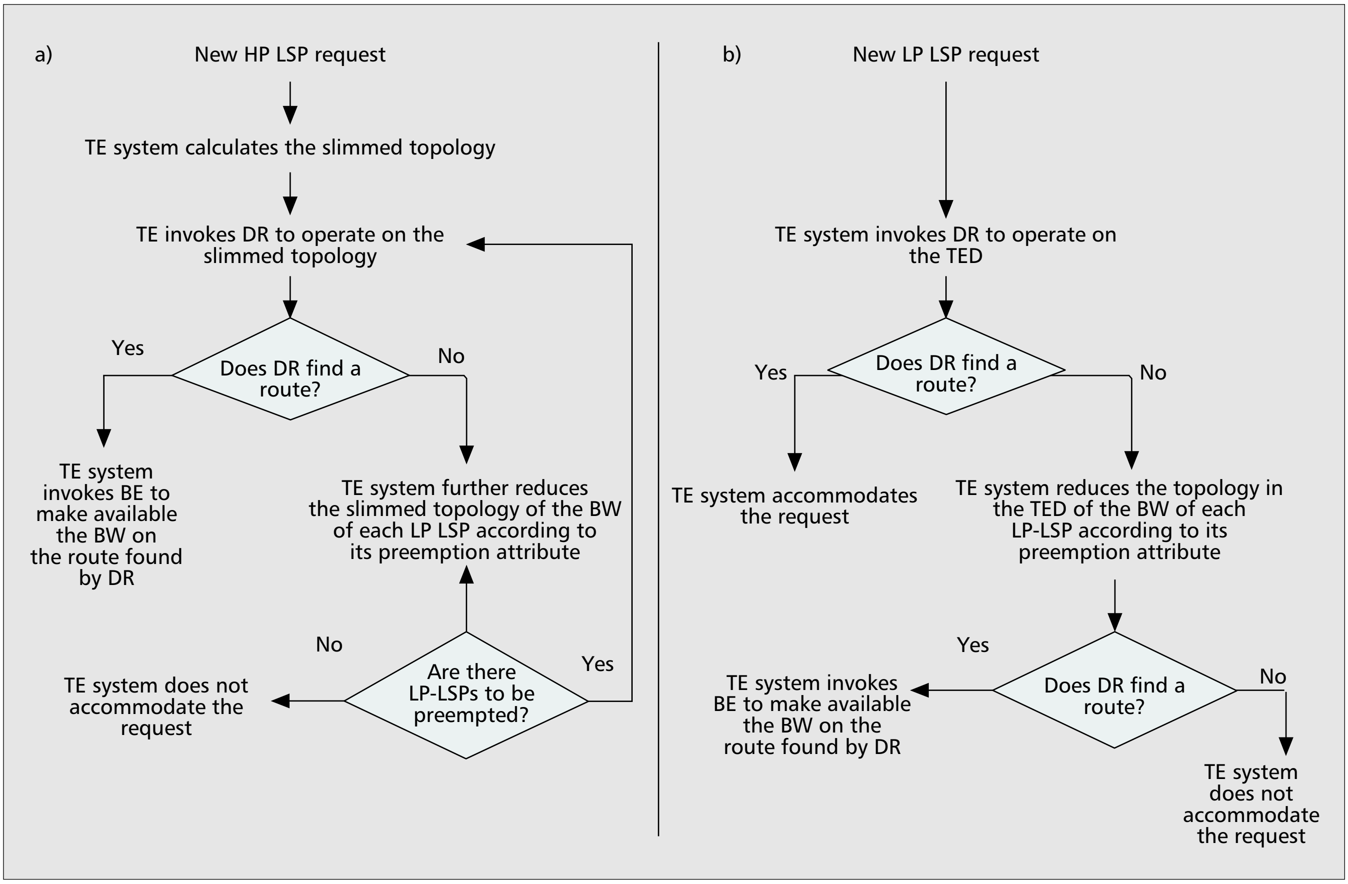
Bandwidth Increase Request — The events relating to bandwidth increase requests are sketched in Fig. 6. The TE checks if the LSP requesting more bandwidth belongs to the HP group (a) or not (b).

a) In the first case, it verifies if the requested bandwidth does not exceed the amount of bandwidth specified in the SLA. If the request does not respect the agreement, it is rejected by the TE system; otherwise, the TE system achieves the modify operation and checks if there is any congestion in any of the links crossed by that LSP. If there is no congestion at all, the TE accomplishes the bandwidth increase and updates the relevant databases. Otherwise, it invokes the BE module, which removes some lower-priority LSPs sharing one or more links of the considered route in order to make available the desired portion of bandwidth for the HP LSP. Finally, the TE system performs the bandwidth increase and updates the databases. In the meantime, the BE module will try to reroute the removed LSPs toward less congested routes. The BE module will tear down the LSPs it did not succeed in rerouting.

b) If the requesting LSP does not belong to the HP group of LSPs, the TE achieves the modify operation. If there is not enough available bandwidth to fulfill the new request, the TE system invokes the BE module as in the previous case. In this case it is not assured that the BE is able to accommodate the request. If unsuccessful the TE system will reject the request. Note that in this way the TE system makes use of all the available network resources in an attempt to accommodate most of the connection requests while honoring the QoS agreements. Any LSP pays the bandwidth it consumes for a certain amount of time, thus realizing bandwidth-on-demand service.


a)
BW increase requested by HP-LSP
TE system verifies if the request meets SLA and if there is enough BW to accommodate the request
Requested BW ≤ MB?
Yes
No
TE system checks if there is enough BW to accommodate the request
TE does not accommodate the request
Is there available BW on each link on the route?
No
TE system iinvokes BE to make available the BW
Yes
TE system accommodates the request
b)
BW increase requested by LP-LSP
TE system checks if there is enough BW to accommodate the request
Is there available BW on each link on the route?
No
TE system invokes BE to make available the BW
Yes
TE system accommodates the request
Yes
Does BE make available the requested BW?
No
TE does not accommodate the request

■ Figure 6. *Workflow of the TE system operation in response to bandwidth increasing requests, relating to a) HP LSPs; b) LP LSPs.*



a)
New HP LSP request
TE system calculates the slimmed topology
TE invokes DR to operate on the slimmed topology
Does DR find a route?
Yes
No
TE system invokes BE to make available the BW on the route found by DR
TE system further reduces the slimmed topology of the BW of each LP LSP according to its preemption attribute
Are there LP-LSPs to be preempted?
No
Yes
TE system does not accommodate the request
b)
New LP LSP request
TE system invokes DR to operate on the TED
Does DR find a route?
Yes
No
TE system accommodates the request
TE system reduces the topology in the TED of the BW of each LP-LSP according to its preemption attribute
Does DR find a route?
Yes
No
TE system invokes BE to make available the BW on the route found by DR
TE system does not accommodate the request

■ Figure 7. *Workflow of the TE system operation in response to new connection requests, relating to a) HP LSPs; b) LP LSPs.*

Connection Request for a New HP LSP — This event is shown in Fig. 7a. When a new HP LSP not predicted by the traffic matrix during the provisioning phase has to be accommodated, the operator can decide to try to accommodate the new request without providing global optimization by means of a new provisioning phase. In this case, the operator can verify the possibility of accommodatiing the new request provided that the new HP LSP does not compete with the other HP LSPs already foreseen and has as little impact as possible on the LP LSPs accommodated in the network. This can be done by calculating the new route with the dynamic routing algorithm on a slimmed network topology. Such a slimmed topology can be obtained by taking the current topology (recorded in the TED) and lowering on the links the amount of bandwidth corresponding to the maximum value of already existing HP LSPs. If the new route is found, the new connection request is accepted; otherwise, the topology is further modified by increasing the amount of bandwidth on the links that is needed to accommodate the new request, by assuming to preempt one or more lower-priority LSPs. Such a procedure is iterated until the DR finds a new route, or when, even preempting all possible LSPs belonging to classes lower than that of the new LSP, the DR cannot find any solution. At the end, if the route is found, the BE is finally invoked to work on that found route and performs its function to actually rereoute or even preempt lower-priority LSPs.

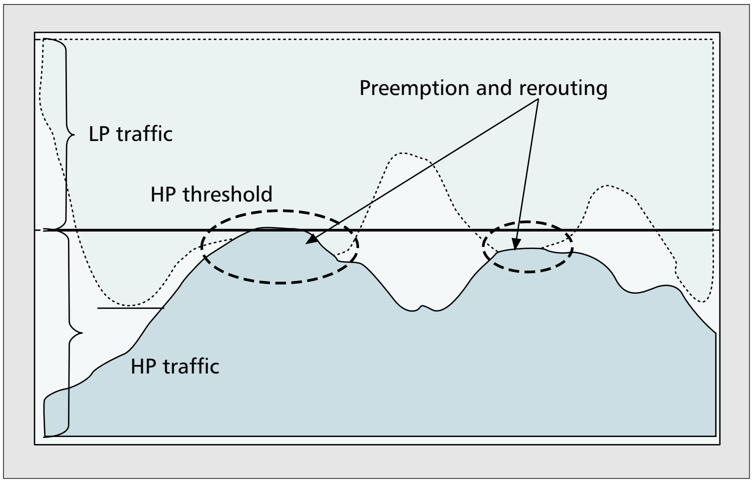
Connection Request for a New LP LSP — This event is sketched in Fig. 7b. This is a common event. The TE invokes the DR and tries to accommodate the new request. If it does not succeed, the same procedure described in the previous point applies, except for the initial operation that slims the topology.

Characteristics of the Proposed Traffic Engineering System

The proposed strategy presents several advantages, considering both performance improvements, with respect to conventional IP/MPLS systems in terms of traffic accommodated while guaranteeing QoS requirements, and feasibility.

The performance improvements basically lay in two key aspects of the proposed solution: the features of the hybrid routing solution, and the realization of the elastic bandwidth concept.

The hybrid routing solution benefits from the advantages of both offline and online procedures. In fact, the path provisioning achieved offline allows the best use of network resources to be attained for all cases in which the traffic can be reasonably predicted; while the dynamic routing function, performed online, provides a prompt reaction anytime it is required to route or reroute LSPs within the multilayer network. In particular, the fact that the network control is aware of all the network elements and is able to manage the whole set of resources is fully exploited by the multilayer PR module to compute routes in an optimal way [19]. On the other hand, the knowledge of the actual status of the network as it changes with time is exploited by the DR for finding routes on the basis of individual requests made on demand with the bandwidth constraint [20]. Clearly, the performance of the DR is based on its knowledge of network status. A detailed and updated network status requires a considerable information flood to be disseminated throughout the network by signaling. We demonstrate that the performance of such a hybrid approach is convincing even in cases where traffic demand changes appreciably with respect to the original traffic matrix [24]. In fact, the simplicity and robustness of the DR algo-



■ Figure 8. *The concept of bandwidth elasticity.*

rithm allows the data flows to be suitably distributed throughout the network, preventing congestion even in critical situations.

The concept of bandwidth elasticity is illustrated in Fig. 8. The basic idea is to make available the portion of bandwidth temporarily released by high-priority traffic in order to accommodate other requests with lower priorities. Actually, the TE system does not waste bandwidth, since the HP traffic occupies only the amount of bandwidth it really needs in a certain period, and the temporarily released bandwidth is put at the disposal of all the other LP services. At the same time, TE does ensure the required bandwidth for all the HP services, by immediately giving back the desired amount of bandwidth on specific request. This is made possible by the BE concept that not only achieves preemption mechanisms to free the bandwidth according to a proper priority policy, but is able to rearrange traffic flows by means of intelligent rerouting.

Altogether, the proposed TE solution favorably applies in a scenario where the actual traffic entering the network changes with time and is not completely predictable. It allows a large amount of traffic to be accommodated with respect to traditional methods based on overprovisioning, while guaranteeing the desired bandwidth attribute for premium connections at any time, whatever the traffic demand.

A basic realization of the TE solution could also make use of simpler building blocks (e.g., using a single-layer approach). In fact, due to the modular structure of the TE system, each building block can evolve almost independently and can be upgraded in order to contribute to the improvement of the overall performance of the system without affecting the applicability of the solution.

A key issue is the practical feasibility of the proposed solution. It uses the known framework of the MPLS control plane and its extensions, which utilize updated versions of well assessed Internet protocols. Specifically, the proposed TE employs key MPLS functionalities such as explicit routing, modification, and preemption, and provides the means to perform CBR functions automatically via suitable signaling and routing protocols. These functions are consolidated by the progress of standardization activities in different bodies, such as IETF, Optical Interworking Forum (OIF), and International Telecommunication Union — Telecommunication Standardization Sector (ITU-T), where the GMPLS model and relevant network interfaces are going to be fixed. While the network paradigm is quite consolidated, it is not yet clear how these functions must be implemented. In the following some relevant issues related to the realization of the TE solution are given, without going into implementation details.

One of the most significant impacts of the convergence of datacom and telecom relates to network control and management (NC&M) functions [26]. These have a strong impact on the way the control is structured: the datacom world pushes for a distributed approach, while the telecom world favors a centralized one. In particular, the use of signaling/routing protocols coming from the datacom world allows automation of some TE operations, such as path setup. The typical requirements of the telecom world claim a certain efficiency of TE operations. A key issue relates to the type of information that needs to be flooded and the frequency of information updating, as discussed earlier. In the following the realization of the proposed TE solution is discussed. It could be useful to distinguish among TE operations that can be performed offline, such as HP route calculation, global path optimization, and lightpath setup/teardown; and TE operations that are performed online, like LP route calculation for LSP setup/rerouting, and preemption. The former type of operations can easily be achieved in both distributed and centralized ways, since the information stored in the databases, as described earlier, is available and updated, so there are no stringent requirements to be met in terms of speed or delay. More relevant is the application of the online TE operations (i.e., CBR and preemption). Here the main issue relates to the specific strategy for preemption. In fact, as described in previous sections, CBR operations just need the information stored in both the RDB and TED. The former is updated offline and the latter dynamically by routing protocol flooding. Differently, preemption needs the DLD database, where the LSP map is stored, besides the RDB and TED databases. Assuming the current status of the MPLS protocol suite defined by the standards, the network control is not able to learn the LSP map for the entire network; each node can know only the LSPs it manages (i.e., the node maintains information just for those LSPs that originate, end, and transit through it). Therefore, each node can preempt only the LSPs it controls. This could lead to nonoptimal choices, with respect to either each node having knowledge of all the LSPs throughout the network (but this leads to modifying the standards somewhere), or a management entity knowing the status of all LSPs throughout the network. Clearly the application of the TE solution assuming centralized control is straightforward. As a result, even if addressing the issue of the choice between a centralized or distributed approach is beyond the scope of this article, a relevant fact is that the proposed TE system can be achieved in both ways, even if with different features and performance. In particular, we demonstrated the feasibility of the distributed approach in [27], where experiments on a real testbed are reported.

Furthermore, it is to be highlighted that the TE system can be utilized in different segments of the network: from the edge of the core network to the metro area, up to the backbone network. What changes in such network segments is the traffic aggregation volumes and the dynamic variability of the traffic itself. In fact, while the backbone traffic is more stable and aggregated on larger data flows, in the edge of the core network the data flows are much smaller and more variable with time. What changes in those cases is the balance between utilization of offline and online routing.

Perspectives and Conclusions

Traffic engineering will be the key feature for the realization of flexible networks able to make effective use of network resources and provide bandwidth-on-demand services. This capability will characterize new-generation network infrastructures required to support different types of services, with several levels of quality of service. While the main paradigm related to the GMPLS control plane is well assessed, as witnessed by the progress of work within the standardization bodies, the architectural aspects that will allow the advanced concepts of traffic engineering to be concretely achieved represent a still open issue. In fact, the realization of effective constraint-based routing algorithms, preemption and rerouting approaches, and adequate signaling is still debated.

The article reports a possible strategy to practically implement traffic engineering in multilayer networks taking advantage of GMPLS control plane features. Such a solution is based on a combined use of offline and online routing, and a novel approach to guaranteeing QoS, preventing congestions, and effectively handling preemption and rerouting of data flows.

Acknowledgments

This article reports a traffic engineering strategy partially developed in the framework of the Ericsson project Asterix. The authors wish to thank all participants in that project. In particular, we are grateful to all the people who contributed to the realization of the key building blocks: Alessandro Bosco and Eleonora Manconi, for the development of the bandwidth engineering concept and its realization; Giulia Conte, for achievement of the path provisioning module; and Yun Xiang Pietro Tou, for the development of the dynamic routing module. We also thank Roberto Mameli for his contribution in developing a testbed for the demonstration of the proposed approach, and Alberto Bricca for fruitful discussions. The authors wish to thank the Ericsson business unit on transport networks for sponsoring the Asterix project. Finally, the authors wish to thank the anonymous reviewers for their precious suggestions.

References

[1] M. Listanti *et al.*, "Architectural and Technological Issues for Future Optical Internet Networks," *IEEE Commun. Mag.,* vol. 38, no. 9, 2000, pp. 82–92.

[2] X. Xiao *et al.*, "Internet QoS: A Big Picture," *IEEE Network*, Mar./Apr. 1999, pp. 8–33.

[3] X. Xiao *et al.*, "Traffic Engineering with MPLS in the Internet," *IEEE Network*, Mar./Apr. 2000, pp. 28–33.

[4] A. Banerjee *et al.*, "Generalized Multiprotocol Label Switching: An Overview of Routing and Management Enhancements," *IEEE Commun. Mag.*, Jan. 2001, pp. 144–50.

[5] G. Ash, "Traffic Engineering and QoS methods for IP-, ATM-, and TDM-Based Multiservice Network" Internet draft <draft-ietf-tewg-qos-routing-04>, Oct. 2001.

[6] D. Awduche et al., "Requirements for Traffic Engineering over MPLS," IETF RFC 2702, Sept. 1999.

[7] E. Rosen *et al.*, "Multiprotocol Label Switching Architecture," IETF RFC 3031, Jan. 2001.

[8] D. Awduche *et al.*, "Overview and Principles of Internet Traffic Engineering," IETF RFC 3272, May 2002

[9] R. Coltun, "The OSPF Opaque LSA Option," RFC 2370, July 1998.

[10] D. Katz, "Traffic Engineering Extensions to OSPF," work in progress, IETF draft <draft-kats-yeung-ospf-traffic-06.txt>.

[11] B. Davie *et al.*, *MPLS Technology and Applications*, Academic Press, 2000.

[12] D. Awduche *et al.*, "RSVP-TE: Extensions to RSVP for LSP Tunnels," RFC 3209, Dec. 2001.

[13] E. Mannie *et al.*, "Generalized Multi-Protocol Label Switching (GMPLS) ,Architecture" work in progress IETF draft <draft-ietf-ccamp-gmpls-architecture-02.txt>.

[14] A. Banerjee *et al.*, "Generalized Multiprotocol Label Switching: An Overview of Signaling Enhancements and Recovery Techniques," *IEEE Commun. Mag.*, July 2001, pp. 144–51.

[15] K. Kompella *et al.*, "LSP Hierarchy with Generalized MPLS TE," work in progress, IETF draft <draft-ietf-mpls-lsp-hierarchy-07.txt>.

[16] OIF Architecture, OAM&P, PLL, & Signaling Working Groups, "User Network Interface (UNI) 1.0 Signaling Specification," OIF200.125.3.

[17] R. Dutta *et al.*, "Traffic Grooming in WDM Networks: Past and Future," *IEEE Network*, Nov./Dec. 2002, pp. 46–56.

[18] J. Wang *et al.*, " Improved Approaches for Cost-Effective Traffic Grooming in WDM Ring Networks: ILP Formulations and Single-Hop and Multihop Connections," *IEEE J. Lightwave Tech.*, vol. 19, no. 11, 2001, pp. 1645–53.

[19] M. Settembre *et al.*, "A Multi-Layer Solution for Paths Provisioning in New Generation Optical/MPLS Networks" accepted for publication, *IEEE J. Lightwave Tech.*

[20] R. Sabella *et al.*, "Strategy for Dynamic Multi-Layer Routing in New Generation Optical Networks based on the GMPLS Paradigm," submitted to *IEEE J. Lightwave Tech.*

[21] P. Ashwood *et al.*, " Generalized MPLS-Signaling functional description," work in progress, IETF draft, <draft-ietf-mpls-generalized-signaling-08.txt>, Oct. 2002.

[22] J. P. Lang, "Link Management Protocol," work in progress, IETF draft, <draft-ietf-ccamp-lmp-04.txt>, May 2002.

[23] A. Shaikh *et al.*, "Evaluating the Overheads of Source-Directed Quality of Service Routing," *Int'l. Conf. Net. Protocols*, 1998.

[24] G. Conte *et al.*, " Off-line and On-line Traffic Engineering Approaches: A Hybrid Solution for Paths Provisioning in Optical/MPLS Networks," *Proc. 7th IFIP Wkg. Conf. Opt. Net. Design & Modeling,* Feb. 3–5, 2003, Budapest, Hungary.

[25] L. Valentini, "Soluzioni Avanzate di Ingegneria del Traffico in Reti Ottiche di Nuova Generazione basate sul Paradigma GMPLS: Strategie ed Analisi Prestazionale," Master thesis, University of Perugia, 2002, to be submitted to *IEEE Commun. Letters.*

[26] O. Gerstel, "Optical Layer Signaling: How Much Is Really Needed?" *IEEE Commun. Mag.*, Oct. 2000, pp. 154–60.

[27] A. Bosco *et al.*, "Distributed Implementation of a Pre-emption and Re-routing Mechanisms for a Network Control Based on IP/MPLS Paradigm," *Proc. 7th IFIP Wkg. Conf. Opt. Net. Design & Modelling,* Feb. 3–5, 2003, Budapest, Hungary.

Biographies

PAOLA IOVANNA (paola.iovanna@eri.ericsson.se) received a Laurea degree in electronics engineering from the University of Roma "Tor Vergata" in 1996. From 1995 to 1997 she had collaboration as a fellow with Fondazione Ugo Bordoni in Rome, where she dealt with advanced fiber optic communications and optical networking issues. From 1997 to 2000 she worked at Telecom Italia, where she was involved in experimentation of new services based on different access technologies (xDSL, frame relay, optical, etc.). In 2000 she joined Ericsson Lab Italy in the research department where she dealt with networking issues using MPLS and GMPLS techniques. She is responsible for a research project relating to traffic engineering strategies based on the MPLS control plane, within a national research project on new-generation networks. She holds a patent on dynamic routing solutions for networks based on the GMPLS model. She has given courses on IP networking, MPLS, GMPLS, and networking issues in Ericsson.

MARINA SETTEMBRE (marina.settembre@eri.ericsson.se) received a Laurea degree in physics from the University of Rome "La Sapienza" in 1985. In the same year she was granted a fellowship at the Fondazione Ugo Bordoni, working on innovative materials for optical devices. From 1986 to September 2000 she worked as a research scientist at the Fondazione Ugo Bordoni in the Optical Communication Division focusing first on optical devices for signal routing/processing and then on physical layer modeling and high-capacity optical transmission systems. She was actively involved in several European COST, ACTS, and IST projects as a work package leader (ACTS-Upgrade, ACTS-Esther, Cost 266, IST Atlas). Since October 2000 she has been with Ericsson Lab Italy working on optical networking in the Research Unit. Her present research interests include network architectures, traffic engineering strategies and related algorithms, and control plane definition for new-generation networks based on the GMPLS paradigm. She is also involved in research activity on broadband wireless access networks. She published more than 90 papers in international scientific journals and conferences, and a book, *Nonlinear Optical Communication Networks* (Wiley, 1998).

ROBERTO SABELLA [SM] (roberto.sabella@eri.ericsson.se) received a Laurea degree in electronic engineering (Laurea in Ingegneria elettronica) in 1987. He then joined Ericsson, Rome, Italy, where he was involved first in hardware design and subsequently in research activities on advanced fiber optic communication systems. His research interests have covered the fields of optical device technology, high-speed optical communication systems, WDM optical networks, and new-generation Internet networks. In May 1997 he became the technical coordinator of the research consortium CoRiTeL. Since 1999 he has been the manager of the research department of Ericsson Lab Italy. He holds four patents related to optical networks and traffic engineering strategies, is co-author of a book on high-speed optical communications, and author and co-author of about 100 papers in international scientific/technical journals and conferences. He has been a lecturer (professore a contratto) at the University of Rome "Tor Vergata" and the Polytechnic of Bari, both in Italy. He is a member of the editorial board of *Photonic Network Communications.* For the same journal he has operated twice as guest editor for special issues on WDM transport networks, and routing and failure restoration strategies in optical networks. He was also guest editor of a special issue on optical networks for *Computer Network.* He has been guest editor twice for IEEE magazines, for the special issues on Optical Networking Solutions for Next Generation Internet Networks of *IEEE Communication Magazine*, and Traffic Engineering in Optical Networks in *IEEE Network.*