Available online at www.sciencedirect.com
SCIENCE DIRECT®

Performance Evaluation 59 (2005) 121–136

PERFORMANCE EVALUATION
An International Journal

www.elsevier.com/locate/peva

Internet like control for MPLS based traffic engineering: performance evaluation

A. Bosco[a], A. Botta[b], G. Conte[b], P. Iovanna[a], R. Sabella[a], S. Salsano[c,*]

[a] *Ericsson Lab Italy, Research and Innovation Unit, Via Anagnina 203, 00040 Rome, Italy*
[b] *CoRiTel—Consorzio di Ricerca sulle Telecomunicazioni, Via Anagnina 203, 00040 Rome, Italy*
[c] *DIE—Università di Roma "Tor Vergata", Via del Politecnico 1, 00133 Rome, Italy*

Available online 11 September 2004

Abstract

This paper reports a realization of a traffic engineering (TE) strategy based on a distributed control plane (Internet-like), and a performance analysis based on experiments carried out on the realized test bed and on simulations. The implementation of control plane is based on standard IP protocols, namely OSPF-TE for routing and RSVP-TE for signaling, which have been suitably extended to support the reported TE strategy. Furthermore, a threshold mechanism for limiting the information flood in the network is reported. The performance analysis assesses the validity of the proposed strategy, by reporting simulation results and measurements achieved on the test bed. In particular, the paper shows that the non-ideality due to real signaling flow and related databases updating are reasonable and in-line with real systems.

Keywords: MPLS, GMPLS, Traffic engineering

1. Introduction

A very challenging task for new generation networks (NGN) concerns the definition of a control plane able to combine the simplicity of use and automation of procedures, proper of the Internet world, with the reliability and high performance provided by traditional Telecom management. Multi protocol label switching (MPLS) and its extension, generalized MPLS (GMPLS), are suitable technologies to satisfy such requirements because allow applying efficient Traffic Engineering functionality in IP contest. In

* Corresponding author. Tel.: +39 06 72597450; fax: +39 06 72597435.
E-mail address: stefano.salsano@uniroma2.it (S. Salsano).

0166-5316/$ – see front matter
doi:10.1016/j.peva.2004.07.010

principle, thanks to the extension of IP based routing and signaling protocols, the control plane should be able to perform network operations automatically, providing a level of quality of service (QoS) that is comparable to that one granted by connection oriented technology (e.g. ATM and SDH).

A lot of relevance is given within standardization bodies to define protocols, strategies, mapping criteria among different technologies to define all the ingredients necessary for applying traffic engineering based on the (G)MPLS paradigm in NGN, but a feasible solution that is able to use such ingredients is still not consolidated.

Particularly, the analysis of which TE functions can be implemented in a distributed way, supporting QoS, without affecting both performance and scalability of the solution is a key issue to be addressed.

The scope of this work is evaluating the performance of a TE strategies bases on a simulation tool where the impact of signaling is not considered and a test bed where a distributed implementation of such TE strategy is performed, making use of real signaling (RSVP-TE) and routing (OSPF-TE) protocols. Specifically control plane architecture for the distributed implementation is presented and analyzed.

The performances of such implementation will be evaluated in terms of QoS capability support and scalability issues.

Moreover, the description of main test beds and relating TE architecture reported in literature are also presented and compared with the proposal implementation, in order to give a reference context of the main issues that are considered in this field.

2. TE strategy

In this work, the performance of the TE solution reported in [1] for a MPLS/GMPLS based scenario is analyzed. The reference scenario is based on a multi-layer network composed by MPLS router (LSR) and optical nodes (see Fig. 1). The end-to-end path in the MPLS layer is named label switched path (LSP), while it is named lightpath in the optical layer. The considered TE solution is based on a hybrid routing approach and a bandwidth engineering mechanism [2] in order to optimize the use of resources and guarantee tight QoS support.

Specifically, the hybrid routing allows calculating the routes for LSP requests using a combination of off-line and on-line methods. The off-line technique allows performing a global optimization of the

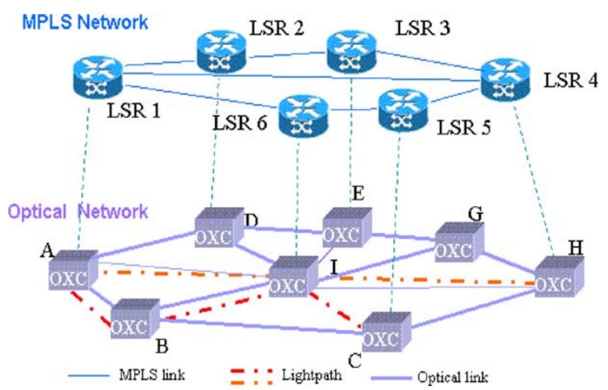

Fig. 1. Reference scenario.

network resources on the multi-layer network. This method is based on the a priori knowledge of the traffic amount that the network has to handle [3]. The on-line approach, on the other hand, allows accommodating each single traffic request that was not predicted, taking into account the actual state of network occupancy, by means of constraint based routing (CBR) described in [4]. Two main classes of traffic are considered: the high priority traffic class (HP) and low priority one (LP). HP class represents the portion of traffic that has to be guarantee whatever is the network load. This means that, at any time and for any traffic distribution in the network, HP traffic must have an amount of available bandwidth up to a maximum value (MV) agreed by service level agreement (SLA).

LP class represents the traffic that is not guaranteed and/or it is not foreseen, thus it can be re-routed, not accommodated or torn down, in case of network congestion.

In each LSP just one type of traffic can be mapped (L-LSP [5]), so, in this work, HP-LSP and LP-LSP are considered. In order to better use the network resources, the bandwidth engineering mechanism is used [2]. This mechanism is based on a pre-emption algorithm and re-routing function as it is explained in the following. Practically, the bandwidth not used by HP-LSP is temporarily released to accommodate LP traffic, due to the use of "modify" function [5]; but, if needed, HP-LSP can require to increase/decrease the amount of bandwidth up to MV, dynamically. In case of bandwidth increasing request, some LP-LSPs could be pre-empted for satisfying the HP request. In order to reduce the loss of LP traffic, the pre-empted LP-LSPs are re-routed in less congested paths.

In the considered TE solution, the multi-layer global optimization is performed by means of a provisioning tool (named PR module), as described in [3].

This module takes as input: (i) the physical topology of the network; (ii) the HP traffic matrix, where the maximum value agreed by SLA is considered; and (iii) an LP expected traffic matrix, whose elements are determined through statistical evaluations. The output of the provisioning module consists of (i) the set of lightpaths (i.e. the optical LSPs according to GMPLS LSP hierarchy) that represents the logical topology of the optical layer, and (ii) the routes for all the LSPs groomed into the lightpaths. Particularly, the routes for HP-LSP are stored in a database named fixed route (FR). The on line route calculation for LP traffic set up/re-routing is performed by means of a dynamic routing algorithm described in [4]. As consequence, both HP and LP traffic are served on demand, but HP traffic routes, off-line calculated, are fixed, while LP traffic routes can be dynamically changed according to the actual network status.

For sake of simplicity, it has been assumed that the dynamic routing module cannot set up new lightpaths during the on-line operations, but it can only operate on the logical topology derived during the off-line provisioning phase. As a result, the establishment of one or more lightpaths can only follow the decision of off-line providing a greater logical capacity to the network due to special events (e.g. when a new Internet service providers enters the network).

This module operates on-line accommodating an LSP at time on the bases of the current status of the network learnt by routing protocol.

The dynamic routing algorithm aims at better utilizing network resources, by using less congested paths instead of shortest, but heavily loaded paths as described in [4], and evenly distributing the traffic at MPLS layer on the logical topology. The bandwidth engineering mechanisms allows handling the connection in flexible way and preventing the congestion when the combination of the hybrid routing is not sufficient. As a result the system allows accommodating also the traffic that was not foreseen during the off-line phase, guaranteeing the QoS for HP traffic. Detailed analysis of the performance of each algorithm that has been used of for each module has been performed in [2–4].

3. Control plane architecture

The TE strategy described in Section 2 can be implemented in different ways: centralized, distributed or hybrid manner. It could be useful to distinguish among TE operations that can be performed "off-line", such as HP route calculation, and global path optimization; and TE operations that are performed on-line, like LP route calculation for LSP set-up/re-routing, pre-emption, and HP modify. The former type of operations can easily be achieved in both distributed and centralized manner, since the information needed to perform such operations can be made available without strict requirements on the signaling, and the level of updating of the databases where the information on the status of the network are stored, is not an key constraint.

More relevant is the application of the "on-line" TE operations: i.e. the CBR and the pre-emption. The efficiency of these functions depends on the updating level of the information stored in the databases described in Section 2. The control plane architecture that has been considered for performing dynamic TE operations using OSPF-TE and RSVP-TE is described in [16]. A challenging target is analyzing the performance of these operations in case of a distributed implementation. The efficiency of routing decisions could suffer from inconsistent co-ordination between the various network elements when the implementation is distributed, due to the time necessary to propagate the information and the imprecision of the information. On the other hand, as it is well known [6], the feasibility of a distributed implementation requires to reduce the amount of signaling that is flooded in the network and meet stability requirements, by means of the use of mechanisms that reduce the level of updating of the information about the network state.

In order to evaluate such features, distributed control plane architecture and its implementation by means of a test bed are considered. Moreover, a comparison with the performance analysis of the TE solution obtained in an ideal environment, without considering real protocols for routing and reservation, is also performed in order to evaluate the performance of the TE solution in two different environments.

4. Testbed implementation

Speaking about traffic engineering realized by means of MPLS functionalities, all the protocol aspects strictly related with the performance obtainable, are the main topics to be analyzed.

As a consequence, it is easy to understand that a prototype implementation of the whole strategy is essential to carry out not only a simulative analysis of different strategies comparison, but also to study consistency aspects, in terms of existing protocols standardization and different protocols interoperability.

The need of an implementation of an MPLS-TE network control plane in a real test bed is even more important because all traffic engineering functionalities are realized by means of protocol extensions and an effective interoperability of each single functional block realizing the node's architecture. Moreover, remarking that all the TE functionalities have to be carried out dynamically in each LSR (distributed architecture), it is clear how important is the possibility to measure the performance reached in terms of rapidity and scalability of all the operations realized by each node.

4.1. Related works

Several TE implementations have been already proposed in literature. Most of them relate with the management plane of MPLS networks. In this sense, a lot of management servers have been developed to monitor network element and traffic performance. In this scenario the RATES server [11] is an important work to remind to. It is a software system developed at Bell Laboratories. It is completely based on centralized paradigm: RATES server communicates with the nodes of the MPLS networks and spawns off signalling from the source to the destination for route LSPs set-up. This kind of communication is viewed as a policy decision and therefore Common Open Policy Service (COPS) protocol is used for this scope. In RATES scenario all traffic engineering aspects are simply treated as routing decision taken by the central server, leading to a leak of dynamicity of the solution proposed.

Another state dependent TE mechanism to distribute network load adaptively is suggested in [12]. MATE assumes that several LSPs have been established between a couple of node in an MPLS domain using a signalling protocol as RSVP-TE. In this way, the ingress node has simply to distribute incoming traffic across the LSPs. It is important to note that MATE is thought for traffic without bandwidth reservation. So, the ingress node has to measure the available bandwidth in term of packet delay and loss and these measurements operations, repeated in each ingress node of the network, could easily lead to loss of dynamicity and scalability. All these problems could be overcome by the introduction of bandwidth reservation and the presence of a link state routing protocol extended with TE capabilities.

TEQUILA [13] is a European collaborative research project looking at an integrated architecture and associated techniques for providing end-to-end QoS in a DiffServ-based Internet. In TEQUILA, an integrated management and control architecture has been designed. The TEQUILA architecture includes control, data and management planes. The management plane aspects are related to the concept of a centralized bandwidth broker (BB). The BB includes components for monitoring, TE, SLS management and policy management. The dynamic route management module considers: (a) setting up the forwarding parameters at the ingress node so that the incoming traffic is routed to LSPs according to the bandwidth determined by network dimensioning, (b) modifying the routing according to feedback received from network monitoring and (c) issuing alarm to network dimensioning in case available capacity cannot be found to accommodate new connection requests.

The control plane aspects are related totally to RSVP signalling protocol and its extensions to support traffic engineering. In this sense, the work carried out in the framework of TEQUILA project is very important, and, among relevant results, it worth to remind the development of an open source version of RSVP-TE protocol. This version of the signalling protocol is equipped with the capability to dynamically setup explicitly routed LSPs with quality of service requirements (only L-LSPs were supported) and represented an important starting point to develop our current version of RSVP-TE used in our implementation.

Another relevant implementation of a traffic engineered MPLS network is represented by the TEAM proposal [14]. TEAM is a traffic engineering automated manager, realized in a central server interacting with the elements of the MPLS network. Various measurements are retrieved from network nodes (queue lengths, available bandwidth, overall delay, jitter, number of packet dropped, etc.) and the network manager takes TE decisions in a completely off-line fashion. All the interfaces to the routers are based on SNMP, COPS and CLI commands. This means that TEAM is related only on management plane aspects and an accurate control plane performance study is still missing.

As a matter of fact, it is evident that low attention has been given to realizing and analysing distributed implementation of the MPLS network's control plane. Moreover, even if several TE strategy have been already proposed, our hybrid approach, resulting in a combination of off-line and on-line TE functionalities, is totally new and unexplored.

4.2. Building blocks

MPLS-TE prototype implementation has been realized by means of general purpose PCs with a Linux Operating System. The reason for choosing this platform is mainly due to the huge availability of open source software and by the possibility to modify and extend existing protocols. The PCs (Pentium III 350 and 600 MHz) are interconnected by means of fast Ethernet point-to-point links.

The starting point for our work was a set of open source software packages:

- MPLS provided by Sourceforge [15];
- RSVP-TE daemon from TEQUILA project [8];
- OSPF-TE daemon by Zebra [7].

With all these pieces running in a stand-alone fashion, there is no integration among all the elements. RSVP-TE and OSPF-TE, for example, run separately, without the possibility to communicate one another. Moreover, there is no “intelligent” functional element like a Route Decision Engine to trigger LSPs set up and tear down.

In order to support the functional architecture described in the previous sections, with all foreseen capabilities, it is necessary to adapt each of the mentioned package to meet the desired behaviour. In this sense, RSVP-TE protocol was extended to support LSP bandwidth modify request, according to the [9]. In addiction, the capability to manage with different classes of traffic was introduced in the RSVP-TE implementation, making it DiffServ aware capable.

Also OSPF-TE protocol implementation was enhanced with an ad hoc mapping of RSVP classes in different Class Types and with the possibility to flood in single Opaque-LSA bandwidth information for all the Class Types. This will lead to a relevant decrement of OSPF message flooding in the network.

In order to further reduce the amount of signalling in the network, a threshold system is also implemented. In particular, such a method reduces the number of OSPF advertisements because a single advertisement is performed after more LSP set ups instead of after each LSP set up.

Several performance study comparisons among different threshold methods are present in literature and a detailed analysis leads to the choice of a “dynamic threshold” mechanism. Such a method considers an initial threshold level on the empty link. The next upper and lower levels are calculated dynamically taking into account the amount of the currently advertised reserved bandwidth (B) and the link capacity (C) according to the following formula:

$$B^{+} = B + F(C - B), \qquad B^{-} = B - F(C - B)$$

where B^{+} is the upper level, B^{-} is the lower level, and $0 \leq F < 1$ is a parameter that regulates the granularity level of the thresholds. The greater F value is the coarser the information flooded by the protocol will be. So, as emphasized in Section 6, OSPF flooding will decrease with higher value of F, with the trade off of not degrade network performances.

To complete the node architecture a “Node Decision Engine” module was implemented. It processes each traffic request incoming to the node, calculate, when needed a constrained shortest path for LSP routing, with the possibility to run different routing algorithms. Furthermore, the Node Decision Engine collects every statistics in terms of request satisfactory and time response of the whole system. In this way it is possible to analyze architectural performances, measuring both the amount of traffic accommodated in the network and the time latency aspects strictly related to protocol behaviour. In Fig. 2 the final node architecture is represented.

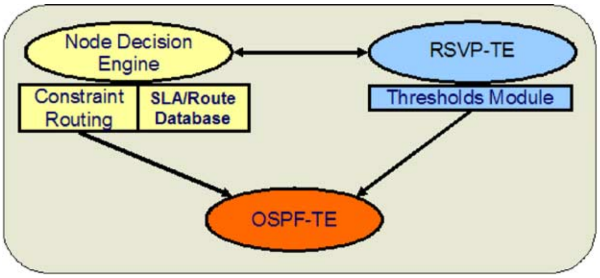

Fig. 2. Node architecture.

4.3. TE-functionalities implementation

The testbed is composed of seven PCs, interconnected according to the topology shown in Fig. 3. The traffic engineered network has been emulated with two classes of traffic, realizing a static mapping of traffic classes (PSCs) into two CTs. Each node is configured with a maximum bandwidth available for each traffic class.

The OSPF-TE distributes the topology and the information of resource utilization per traffic class. It is able to send the unreserved bandwidth values for every CTs in a single opaque LSA, as said in precedence. At the beginning all resources are available. Following the LSP setup procedures, achieved with RSVP-TE signalling, resources are removed along the paths taken by the LSPs. The new bandwidth availability values are communicated through OSPF flooding, with the described mechanisms to control the amount of the exchanged information. In this way, each LER is able to select a constraint based route for every LSP request it receives, with a good approximation of resources occupation in the whole network at the moment the request is received.

For sake of generality, each node can act as an edge node: it can be either source or destination of a given path. When it acts as source, it reads the route of HP-LSPs from the FR, which has been populated according to the provisioning module procedures [3], or calculates the route of LP-LSPs using the dynamic routing algorithm [4].

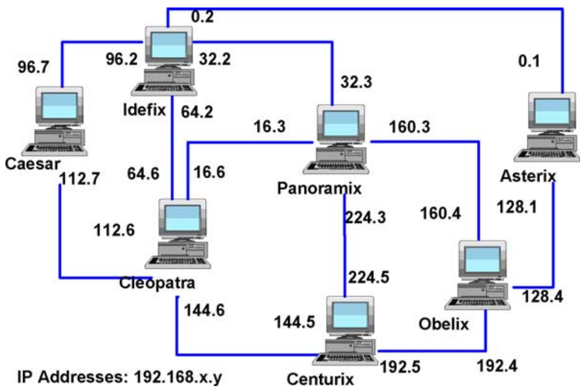

Fig. 3. Testbed topology.

The resource reservation is performed by means of RSVP-TE, using the admission control function that is performed node by node according to the version described in [9].

The architecture emulated in the test bed in its entireness offers the possibility to perform a complete comparison study among different strategies or different algorithms evaluating all the aspects related to the control plane performances.

The evaluation criteria considered here for TE strategy performance are: (i) the amount of traffic accommodated with TE with respect to the one in a traditional over provisioning; (ii) the QoS provided to HP and LP traffic; and (iii) the scalability of the solution that is implemented in distributed way and makes use of routing and signalling protocols to automate network operations. The accommodation of more traffic is realized by means of the bandwidth engineering system [2], which allows an elastic use of the bandwidth of each link. In this way, each request uses just the amount of bandwidth that it needs for the desired amount of time, putting at disposal the unused portion of bandwidth resource when this is not utilized ad the advantage of the lower priority traffic requests. As a result an efficient use of the network resources is obtained.

As far as the QoS the following considerations apply. The TE assures QoS to HP paths by dimensioning their routes at the MV during the off-line optimization, by using bandwidth engineering to prevent possible congestions along the HP routes, and guarantees the agreed bandwidth for HP-LSP whatever is the traffic load in the network. Even if the bandwidth temporarily not used by HP is put at disposal to LP-LSPs, TE strategy has to operate in way that HP-LSP ignores this fact. In other words the HP-LSP bandwidth-increasing request must be always fulfilled without delay even if some LP-LSPs occupy a portion of its bandwidth. It is worth to remind that modify is performed by means of RSVP-TE. Specifically, a local admission control is performed in order to vary the bandwidth attribute of an already set up HP-LSP. Thus, bandwidth engineering (BE) operations have been implemented as explained in the following. Node by node the admission control verifies the availability of bandwidth. In case of lack of resources, the admission control is stopped, the BE algorithm, that is implemented within the admission control function, selects the LP-LSPs to be pre-empted that are pre-empted at once, then the admission control success and the modify operation proceeds on the entire HP path. As far as LP-LSPs is concerned, the TE strategy aims at providing QoS also for LP traffic that is superior to “best effort” performance. Specifically, the selection criterion, described in [2], aims at minimizing the number of LP-LSPs that are pre-empted. In fact, LSPs presented in the link are sorted according their bandwidth attribute, and the minimum number of LSPs that can satisfy the request is chosen. Taking into account that the control is distributed, the selection criterion can utilize just the local information at disposal of that node relating to the LSPs that it handles.

After having selected through BE the LP-LSPs that will be preempted, the node notifies this event to the source node of that LP-LSP, so the source node can promptly operate the re-routing of that LP-LSP.

In order to minimize re-routing time, the pre-emption notify uses RSVP-TE [9]. Specifically, for each pre-empted LP-LSP, the node where the pre-emption has been performed sends two messages concurrently. The first one (“PATH ERROR” with a code that indicates pre-emption event) is sent upstream towards the source node that provides to make free the resources up to the congested link by means of a PATH TEAR message; the second one (“PATH TEAR”) is sent towards the destination node in order to releasing the resources from the node where the pre-emption has been carried out to the destination node at once. Then the source node provides to re-route the pre-empted LP-LSP if it is possible, otherwise that LP-LSP is not accommodated.

In summary, the strategy minimizes the number of LP-LSPs that have to be re-routed, reduce the number of LP-LSPs torn down after the pre-emption owing to network congestion, and minimize the time to perform re-routing. The fulfilment of such requirements allows superior QoS be obtained for LP-LSP with respect to traditional best effort.

5. TE performance analysis

The TE performance analysis has been carried out using both a simulation tool and the test bed described in Section 4. The simulation tool allows analyzing the TE strategy without considering the impact of signaling proper of a distributed control, while the test bed represents a possible solution for implementing such TE strategy in a MPLS based network. Specifically, in the simulation tool, a centralized unit performs TE decisions on the basis of very updated information about the state of the network and the operations on the LSPs, such as set up, tear down, and modify, are performed instantaneously, while in the test bed, as it has been described in Sections 3 and 4, the distributed implementation refers to the on-line operations that are performed on the MPLS layer.

A detailed comparison between centralized and distributed implementation of the strategy is out of the scope if this paper, but it worth to investigate the performance of the TE strategy in a distributed implementation with respect to an ideal case.

5.1. Analysis environment

The traffic for off- and on-line routing is generating according to the following description in both simulation tool and test bed.

The off-line routing is actualized by the global path-provisioning (PR) module that has in input a traffic matrix, composed by HP and LP traffic, whose generic element B_{ij} is the aggregated bandwidth considering the set of LSP requests between node i and node j:

$$[B]^k = \begin{bmatrix} \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & B_{ij}^k & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{bmatrix}, \quad k = \mathrm{HP},\ \mathrm{LP}. \tag{1}$$

The sum of the bandwidth requested by all the LSPs for each pair of nodes is the traffic volume, TV:

$$\mathrm{TV} = \mathrm{TV}^{\mathrm{HP}} + \mathrm{TV}^{\mathrm{LP}} = \sum_{i,j \in V} B_{ij}^{\mathrm{HP}} + \sum_{i,j \in V} B_{ij}^{\mathrm{LP}}. \tag{2}$$

In order to characterize the traffic entering the network, a network load parameter, ρ, defined as the ratio between the total offered bandwidth and the network available bandwidth has been introduced and given by

$$\rho = \frac{\overline{h}\,\mathrm{TV}}{C_{\mathrm{net}}}, \tag{3}$$

where $\overline{h}$ is the average minimum distance between each pair of source-destination nodes, and C_{net} is the total available bandwidth on the physical optical network.

As far as the traffic requests used during the simulation, the aggregated bandwidths are chosen randomly. In particular, for each pair of nodes, *i* and *j*, a random number uniformly distributed between 0 and 1, C_{ij} $\hat{I}[0, 1]$, is picked so that the aggregated bandwidth of all the LSPs from node *i* to node *j*, B_{ij}^k, is $r^{(k)}C_{ij}$, where $r(k)$ is a scaling parameter. The scaling parameter is chosen so that

$$\sum\nolimits_{ij} r^{(k)} C_{ij} = \mathrm{TV}^k, \quad k = \mathrm{HP},\ \mathrm{LP}. \tag{4}$$

According to the proposed TE strategy, the off-line procedure operates using as input the estimated traffic matrix, while the on-line procedures (LP set-up and HP bandwidth modify) operate using as input the actual traffic matrix.

In case of estimated traffic, B_{ijk} represents the average expected aggregated bandwidth from node *i* to node *j*, determined by statistical evaluations, for the LP traffic; while it represents the maximum allowed amount of traffic from node *i* to node *j*, agreed by SLAs, for HP traffic. In the simulations, the estimated traffic matrix has been derived by generating a set of N_{ij} LSP requests with $b_{ij}(n)$, representing the bandwidth associated to each LSP from node *i* to node *j*, so that

$$\sum_{n=1}^{N_{ij}} b_{ij}^{(n)} = B_{ij}^k, \qquad b_{\min}^k \le b_{ij}^{(n)} \le b_{\max}^k, \quad k = \mathrm{HP},\ \mathrm{LP}, \tag{5}$$

where $b_{\min}^k$ and $b_{\max}^k$ represent the minimum and the maximum estimated bandwidth requested by an LSP in case of LP traffic, while they represent the range values defined for the SLAs, for HP traffic.

In case of actual traffic matrix, it has been assumed for LP traffic requests that the connection arrival process between node *i* and node *j* follows a Poisson distribution, with rate λ_{ij} and the connection holding time follows a negative exponential distribution with mean $1/\mu$. The bandwidth of each LP-LSP is uniformly distributed between $b_{\min}^k$ and $b_{\max}^k$, with mean $b = (b_{\max} - b_{\min})/2$. Thus, the average aggregated bandwidth, B_{ij}, between *i* and *j*, can be expressed as follows:

$$B_{ij} = \frac{\lambda_{ij}}{\mu} b. \tag{6}$$

In the simulations, by fixing μ, bmin, and bmax, from formula (6), it is possible to get λ_{ij} for each source-destination pair (i, j), and, hence, to generate the process.

For HP traffic, bandwidth modify events are generated for each HP LSP. The arrival time of bandwidth modify event is assumed to be uniformly distributed between $t_{\min}$ and $t_{\max}$, with mean $t = (t_{\max} - t_{\min})/2$. The amount of bandwidth modify is uniformly distributed between $b_{\min}$ and $b_{ij}(n)$, with mean $b = (b_{ij}(n) - b_{\min})/2$, where $b_{ij}(n)$ is specified in the HP traffic matrix, and represents the SLA for each LSP.

Essentially, the traffic generation bases on the assumption that a traffic matrix, derived by SLAs, is at disposal of the provider for the HP traffic, hence the maximum actual traffic is assumed consistent with the estimated traffic used in the provisioning phase. In the case of LP traffic, instead, the actual traffic can exceed and/or mismatch in spatial distribution the estimated one.

In all the simulations, the value of the holding time, $1/\lambda$, is assumed constant for all the LP-LSP connections and it is 200 s. The LP-LSP bandwidths are assumed to be uniformly distributed from 1 to 500 Mb/s. For the HP traffic, the average modify holding time for each LSP is 2% of the simulation duration. The HP LSP SLAs range between 1 and 500 Mb/s and the modified bandwidth are assumed to be uniformly distributed from zero to the maximum bandwidth allowed by each LSP SLA.

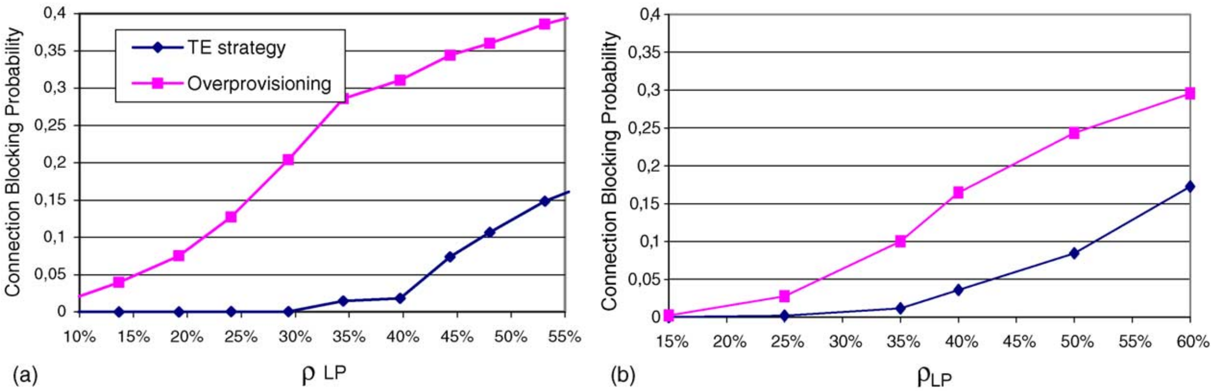

Fig. 4. TE strategy vs. overprovisioning in case of the simulation tool (a) and test bed (b) for different percentage of network load due to LP traffic.

5.2. Performance analysis result

The objective of the performance analysis is twofold: on one hand assess the good TE strategy operations and performance in comparison with conventional systems based on a classic “over-provisioned” approach; on the other hand to demonstrate that the concrete realization of the control plane and its implementation through a test bed allows good performance to be obtained.

For that reason the results reported here were obtained either by a simulation tool, aimed to evaluate the system performance in terms of the blocking probability of required connection (usually named connection blocking probability); or by means of measurements achieved utilizing the test bed.

Several simulations were run in order to analyze the performance of the TE system, in different conditions. For the sake of brevity just a relevant case is reported here, equivalent to the network topology corresponding to the test bed, described in Section 4. Specifically, Fig. 4a and 4b shows the comparison between the proposed TE strategy and an over-provisioning system, in terms of connection blocking probability with respect to the network load due to LP traffic. In particular, Fig. 4a represents the results obtained by means of the simulation tool while Fig. 4b shows the results achieved through measurements on the test bed. It is assumed that HP traffic occupies 50% of the network resources as maximum value agreed by SLA. The connection blocking probability is defined as the number of rejected connection requests with respect to the total number of connection requests. It is worth noticing that both the simulation and experimental results, even though not identical, lead to a common conclusion: the proposed TE strategy leads to reduce the connection blocking probability of about 80% in case of average network load of 40% due to LP (total network load is about 60%), and 100% in case of low network load. The difference between the two figures essentially lays in the fact that the simulation cannot consider all the aspects really affecting the performance as the test bed does.

In the following the analysis focuses two main aspects: (i) the good operations of the TE independently of the implementation, and (ii) the good performance of the implementation solutions that have been achieved.

The first aspect basically depends on the way the hybrid routing works. In fact, thanks to the cooperation between off- and on-line procedure, which aims at even distributing the traffic in the network, the congestion is reached for a value of total network load major of 75%. Practically, the provisioning tool,

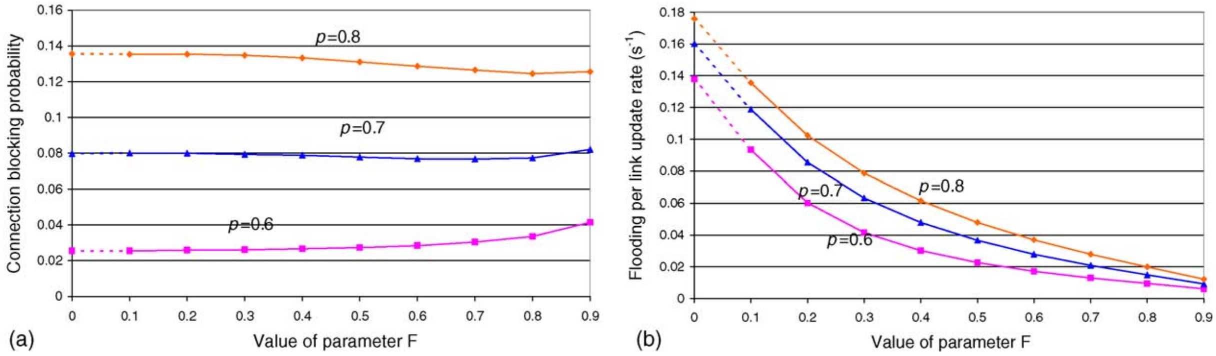

Fig. 5. CBR performance (a) vs. flooding reduction (b) in presence of dynamic threshold mechanism.

that operates in the same way in the cases with and without signaling, aims at evenly distributing the HP routes on the network in order to make available a consistent amount of bandwidth on each link for LP traffic [4], while the dynamic routing, as explained in [4], minimizes the network congestion favoring an even distribution of the LP traffic. Actually, potential congestions that would occur in given portions of network are prevented.

As aforementioned, the test bed allows considering an important aspect that the simulation does not: the impact of non-idealities relating to the signaling process on the performance of both on-line operations (dynamic routing and pre-emption). Such non-idealities include the way the signaling flood throughout the network by means of real protocols is achieved, the real delays due to messages propagation throughout the network, and the not instantaneous updating of the databases. Furthermore, the threshold mechanisms previously introduced to significantly limit the information flooding, contribute to impact the performance.

In order to evaluate the robustness of the routing algorithm with respect to the threshold solution implemented in the test bed that is described in Section 4, in Fig. 5, the performance of the dynamic routing for different network loads is reported varying the factor F. Specifically, when F is 0.7 (Fig. 5a), the connection blocking probability is substantially constant, while the flooding reduction (Fig. 5b) is of about 78% with respect to the situation without threshold.

The other function that has been analyzed is the pre-emption. As it has been mentioned in Section 2, the algorithm that is used to perform the pre-emption aims to minimize the number of LSPs that are pre-empted [2]. Moreover, the implementation of such mechanism, as described in Section 4, is performed in order to meet the following requirements: HP modify is instantaneously, the LP pre-emption notify is very fast and the re-routing of the LP pre-empted traffic is performed with the dynamic routing algorithm previously described, that minimizes the number of connections that are accommodated.

In order to better evaluate the amount of LP traffic that is re-routed and the corresponding percentage that is lost after re-routing, Fig. 6 shows the components of blocking probability of LP traffic. Specifically the component of LP blocking probability due to re-routing range from 0% (LP network load = 15%) to 25% (LP network load = 50%), with respect to the total blocking probability value.

The percentage of LP traffic pre-empted has been evaluated for different HP loads (40, 50 and 60%) as it is shown in Fig. 7. It is worth to highlight that in the worst case, when the maximum value of HP traffic represents 60%, and the average LP traffic is 60% of the total network resources respectively, the percentage of pre-empted LP-LSP, is about 10%.

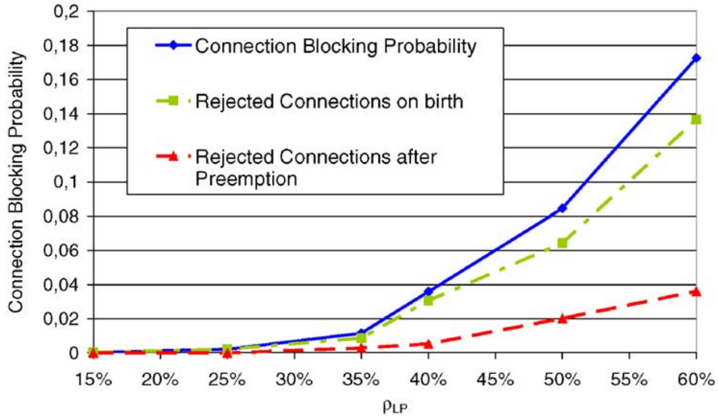

Fig. 6. Blocking probability components.

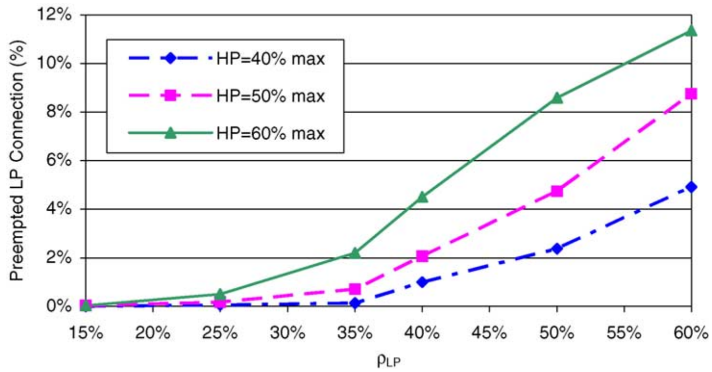

Fig. 7. Preempted LP connections.

Another important parameter to be analyzed is the time necessary to perform HP modify and LP re-routing operations. As it is shown in the Table 1, the average time values for HP set-up and modify coincide, and they are very similar to LP set-up time. This means that HP modify does not experiment delay due to pre-emption and it is slightly dependent of network load. As far as the re-routing time for LP-LSP is concerned, it ranges from 88 ms (60% of total network load) to 112 ms (85% of total network load).

Table 1
Averaged time for set up, re-routing, modify and percentage of preemption per minute

	ρ_{TOT} (%)					
	40	50	60	65	75	85
Preemption per minute	0.03	0.2	0.9	2.6	6	11.1
T_{med} set-up LP (ms)	53.1	59.7	68	76.4	80.5	85.1
T_{med} re-routing LP (ms)	–	–	88.2	95.7	110	112
T_{med} set-up/modify HP (ms)	54.2	60.4	68.1	75.3	79	83.2

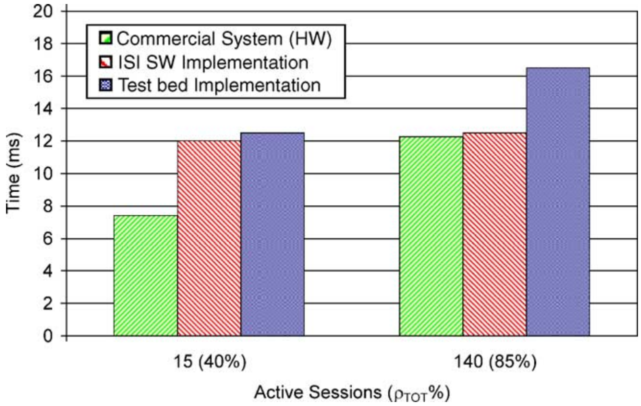

Fig. 8. RSVP propagation time comparison.

With the aim of giving an idea about the processing load for each node due to the pre-emption, the number of pre-emption per minute is also considered (Table 1). Specifically for a network load of 40%, the number of pre-emption per minute is 0.03, while in case of high network load (85%) it is equal to 11.1.

In order to evaluate the goodness of such results, in Fig. 8a, comparison between the measurements collected in the test bed and those ones reported in literature regarding the propagation time for RSVP messages, is performed. In case of 15 active sessions per node (40% of network load), the time measured in the test bed and the value reported in literature is about 12 ms against 8 ms for commercial routers [10], while in case of 140 active sessions per node the measurements performed by the test bed are about 16 ms, while for the other two cases are about 12 ms.

6. Conclusions

In this paper, the performance analysis of a traffic engineering (TE) strategy for MPLS based network, described in [1] is carried out. Specifically the implementation based on a distributed control plane (Internet-like) has been investigated and realized by means of a test bed where real signalling protocol (RSVP-TE) and routing protocols (OSPF-TE) have been implemented. Moreover, the extension of such protocols has been performed in order to support the TE functions.

Specifically, the TE strategy is based on the combination of hybrid routing (off- and on-line methods) and bandwidth engineering (BE) mechanism, that makes use of *modify* and re-routing functions. Two classes of service are considered high priority (HP) class and low priority (LP) class. In order to preserve the feasibility of the solution in real products, standard versions of routing (OSPF-TE) and signalling (RSVP-TE) protocols have been implemented. Scalability issue has been also considered. In fact the implementation of the TE strategy has been performed limiting the OSPF-TE flooding by means of a threshold based mechanism. This mechanism consists in a solution that controls OSPF-TE flooding dynamically in the bases of the occupancy of the link, in order to performs more flooding only when the link is very loaded. The robustness of the routing algorithm that is used for the on-line operations with respect to this mechanism has been also analyzed.

Furthermore, a simulation tool, that is based on a centralized unit that performs routing decision in ideal environment, has been used in order to evaluate the performance of the TE strategy without considering the impact of the signaling proper of a distributed control.

The results obtained using either the simulation tool and the test bed, show that the TE strategy allows to accommodate a superior amount of traffic with respect to a traditional over provisioning where bandwidth overbooking is performed, while providing different QoS. Specifically, in order to evaluate the performance of the TE strategy the following considerations apply. Tight QoS for high priority (HP) traffic is provided, because HP-LSP are guaranteed up to a maximum bandwidth value agreed by SLA, whatever is the network load. The HP traffic uses just the bandwidth that it needs (minor of maximum value) putting at disposal to LP traffic the bandwidth that is not used temporarily, but HP-LSPs do not suffer delay when it requires bandwidth increase if this portion of bandwidth is occupied by LP traffic.

As far as the QoS for LP traffic is concerned, good performances are also obtained, even if the LP-LSP pre-emption algorithm, that minimizes the number of the LP-LSP to be pre-empted, uses just local information of the LP-LSP, due to the fact that the implementation of the solution is distributed. Specifically, the simulation results, obtained by means of the test bed, shown that the percentage of LP-LSPs that are pre-empted is very slight (2% in case of 65% of average network load), and the amount of LP traffic that is torn down after pre-emption is small. In fact, the LP blocking probability due to re-routing is a slight component of the total blocking probability (set-up and re-routing blocking probability). This allows providing a QoS better that traditional best effort also for traffic that is handled dynamically.

Moreover a "*pre-emption notify*" mechanism has been also implemented, in order to perform LP-LSP re-routing quickly. Specifically the average time values for HP set-up and modify coincide, and they are very similar to LP set-up time and are slightly dependent of network load. As far as the re-routing time for LP-LSP is concerned, an interesting result is obtained (re-routing time ranges from 88 to 112 ms for 60 and 85% of the total network load, respectively). The processing load has been also estimated in terms of number of pre-emption per minute that range between 0.03 and 11.1 for 40 and 85% of high traffic load. In order to better evaluate the measurement obtained by the test bed, a comparison between the measurements collected in the test bed and those ones reported in literature regarding the propagation time for RSVP messages, is also performed. The results shown similar results in case of 15 sessions per node (12 ms versus 8 ms of commercial routers), while more difference has been verified in case of 140 sessions per node (16 for the test bed against 12 for the commercial routers).

The next step of this work will be introducing data plane in order to perform specific measurements on the traffic.

References

[1] P. Iovanna, R. Sabella, M. Settembre, Traffic engineering strategy for multi-layer networks based on the GMPLS paradigm, IEEE Netw. 17 (2) (2003) 28–37.

[2] A. Bosco, E. Manconi, R. Sabella, L. Valentini, An innovative solution for dynamic bandwidth engineering in MPLS networks with QoS support, Photonic Netw. Commun. 7 (1) (2004) 37–42.

[3] M. Settembre, R. Sabella, G. Oriolo, F. Razza, F. Ferlito, G. Conte, A multi-layer solution for path provisioning in new generation optical/MPLS networks, J. Lightwave Technol. 21 (5) (2003) 1141–1155.

[4] R. Sabella, P. Iovanna, G. Oriolo, P. D'Aprile, Strategy for dynamic routing and grooming of data flows into lightpaths in new generation networks based on the GMPLS paradigm, Photonic Netw. Commun. 7 (2) (2004) 131–144.

[5] D. Adwuche et al., SVP-TE: extensions to RSVP for LSP tunnels, IETF RFC 3209.

[6] A. Shaini, et al., Evaluating the overheads of source direct quality of service routing, in: Int. Conf. Netw. Protocols (ICNP), 1998.
[7] Zebra Home Page, http://www.zebra.org/.
[8] IBCN test lab Home Page http://dsmpls.atlantis.rug.ac.be/.
[9] L. Berger et al., Generalized multi-protocol label switching (GMPLS) signaling resource ReserVation protocol-traffic engineering (RSVP-TE) e, IETF RFC 3473, proposed standard.
[10] K. Németh, G. Fehér, I. Cselényi, Benchmarking of signaling based resource reservation in Internet, in: Networking, 2000.
[11] P. Aukia, M. Kodialam, P.V.N. Koppol, T.V. Lakshman, H. Sarin, B. Suter, RATES: a server for MPLS traffic engineering, IEEE Netw. 14 (2) (2000) 34–41.
[12] A. Elwalid, J. Ching, S. Low, I. Widjaja, MATE: MPLS adaptive traffic engineering, in: Proceedings of IEEE INFOCOM'01, Anchorage, USA, 2001.
[13] P. Trimintzios, L. Georgiadis, G. Pavlou, D. Griffin, C.F. Cavalcanti, P. Georgatsos, C. Jacquenet, Engineering the multi-service Internet: MPLS and IP-based techniques, in: Proceedings of IEEE International Conference on Telecommunications, Bucharest, Romania, 2001.
[14] I.F. Akyildiz, T. Anjali, L. Chen, J.C. de Oliveira, C. Scoglio, A. Sciuto, J.A. Smith, G. Uhl, A new traffic engineering manager for DiffServ/MPLS networks: design and implementation on an IP QoS Testbed, Comput. Commun. 26 (2003) 388–403.
[15] Source forge MPLS home page, http://mpls-linux.sourceforge.net/.
[16] A. Botta, P. Iovanna, R. Mameli, G. Piantanida, S. Salsano, Extending MPLS traffic engineering to deal with QoS, in: IEEE Softcom 2003, 7–10 October 2003. Split, Dubrovnik (Croatia), Ancona, Venice (Italy).