

A Hierarchical Multiobjective Routing Model for MPLS Networks with Two Service Classes

José Craveirinha, Rita Girão-Silva, João Clímaco, and Lúcia Martins

Abstract This work presents a model for multiobjective routing in MPLS networks formulated within a hierarchical network-wide optimization framework, with two classes of services, namely QoS and Best Effort (BE) services. The routing model uses alternative routing and hierarchical optimization with two optimization levels, including fairness objectives. Another feature of the model is the use of an approximate stochastic representation of the traffic flows in the network, based on the concept of effective bandwidth. The theoretical foundations of a heuristic strategy for finding “good” compromise solutions to the very complex bi-level routing optimization problem, based on a conjecture concerning the definition of marginal implied costs for QoS flows and BE flows, will be described. The main features of a first version of this heuristic based on a bi-objective shortest path model and some preliminary results for a benchmark network will also be revealed.

José Craveirinha e-mail: jcrav@deec.uc.pt · Rita Girão-Silva e-mail: rita@deec.uc.pt
Department of Electrical Engineering Science and Computers of the University of Coimbra Pólo II, Pinhal de Marrocos, P-3030-290 Coimbra, Portugal Institute of Computers and Systems Engineering of Coimbra (INESC-Coimbra) R. Antero de Quental, 199, P-3000-033 Coimbra, Portugal

João Clímaco
Faculty of Economics of the University of Coimbra Av. Dias da Silva, 165, P-3004-512 Coimbra, Portugal Institute of Computers and Systems Engineering of Coimbra (INESC-Coimbra) R. Antero de Quental, 199, P-3000-033 Coimbra, Portugal e-mail: jclimaco@inescc.pt

Lúcia Martins
Department of Electrical Engineering Science and Computers of the University of Coimbra Pólo II, Pinhal de Marrocos, P-3030-290 Coimbra, Portugal Institute of Computers and Systems Engineering of Coimbra (INESC-Coimbra) R. Antero de Quental, 199, P-3000-033 Coimbra, Portugal e-mail: lucia@deec.uc.pt

Work partially supported by programme POSI of the EC programme cosponsored by FEDER and national funds.

1 Introduction and Motivation

Modern multiservice network routing functionalities have to deal with multiple, heterogeneous and multifaceted QoS (Quality of Service) requirements. This led to routing models, the aim of which is the calculation and selection of one (or more) sequences of network resources (designated as routes, which correspond to loopless paths in the network representation) satisfying certain QoS constraints and the optimization of route related metrics. Therefore there are potential advantages in formulating important routing problems in these types of networks as multiple objective optimization problems. These formulations enable the trade-offs among distinct performance metrics and other network cost function(s) to be pursued in a consistent manner. Note that the definition of the objective functions and constraints depends strongly on the nature of the considered routing principles, the type of network technological platform and the features of the offered traffic flows associated with different service types.

In these networks, connection-oriented services, namely with guaranteed QoS levels, may be implemented. The traffic flows are composed (at the physical level) of packet streams that are forwarded from node to node, according to some specific technical rules. When the packets enter the network, they are grouped in different FECs (Forward Equivalence Classes) according to specific criteria, namely the originating node, the destination node and the grade of service that has to be provided. The ‘traffic trunks’ are an aggregation of flows of a certain class and can be characterized by the ingress and egress nodes, the FEC they are associated with, and a set of parameters/attributes with impact on the traffic engineering schemes, which define some essential requirements of the routing models. The routing mechanism for packets used in the MPLS networks is based on the establishment of the so-called LSPs (Label Switched Paths). At the ingress node, a label containing information on the FEC is associated with the packets. At each intermediate node, the LSRs (Label Switching Routers) forward the packets using a specific label switching technique: the label is an index into a routing table with information on the next network arc (or hop) and the next label to be assigned to the packet. Therefore, end-to-end “explicit routes” may be established in association with the implementation of advanced QoS-based routing mechanisms. In particular explicit routes enable source routing mechanisms, characterized by the fact that the route followed by each packet stream (of a given connection) is entirely determined by the ingress router. This is an inherent advantage by comparison with the hop-by-hop (i.e. node by node) routing method typical of IP routing. Details on traffic engineering-related concepts in MPLS networks relevant in the present context are described in [2, 3, 33]. The described features in association with other functional capabilities of MPLS enable the implementation of advanced QoS-based routing mechanisms, namely through the definition of explicit routes satisfying certain QoS requirements for each traffic flow of a given FEC.

Having in mind these features and capabilities of MPLS routing a significant number of routing models has been proposed in the literature in recent years. A routing model can be described in terms of various features. A key feature is the

routing optimization framework which has to do with the scope and nature of the formulation of the routing calculation problem; in this respect we may distinguish network-wide optimization models and flow-oriented optimization models. In the former the objective functions are formulated at network level and depend explicitly on all traffic flows in the network. Examples of these functions are average total traffic carried, total expected revenue, average packet delay or a function which seeks the optimization of the utilization of the arcs of the network in terms of their level of occupancy, as in [13] and [26]. In contrast, flow-oriented optimization models consider the objective functions formulated at the level of each particular node-to-node connection or flow, for example number of arcs of the path, path cost (for a specific link usage path metric) or mean packet delay on the particular traffic stream. Examples of this type of models are the numerous QoS routing models which are based on single-objective constrained shortest path formulations (a review can be seen in [22] and an overview in [4]). Another feature refers to the *nature* of the chosen objective functions and constraints, namely whether the optimization model is single or multiobjective, and the type of functions and constraints (technical, economic, social or other). The *representation* of node-to-node demand requests or traffic offered is also relevant in a telecommunication routing model. We can consider different types of traffic models in terms of the granularity of the representation (for example representation at connection request level or traffic flow level i.e. in terms of a sequence of connections throughout time), or the nature of this representation, namely whether it is deterministic or stochastic.

A recent systematic, in-depth review on main issues, optimization models and algorithms concerning routing methods in communication networks can be seen in [31]. An overview of some applications of MCDA (Multicriteria Decision Analysis) tools in telecommunication strategic planning and negotiation is shown in [16]. A review on applications of MCDA in telecommunications network planning and design, including a section on routing models, is presented in [4]. An overview of a significant number of contributions on multicriteria routing models in telecommunication networks followed by a description of a bi-level hierarchical multicriteria routing model of the flow-oriented optimization type, is put forward in [5].

As discussed in [7], a significant number of routing models for MPLS has been proposed in the literature in recent years which often differ in key instances of the modeling framework. Based on the analysis of the remarkable differences observed in the models proposed in this area, a discussion on key conceptual issues involved in the various modeling approaches and a proposal of a generic hierarchical multiobjective network-wide routing optimization framework or “meta-model” has been presented in [7].

The possibility of applying this modeling framework to a MPLS type network, already outlined in [7], namely by considering two classes of service, QoS traffic (first priority traffic) and Best Effort (BE) traffic (second priority traffic), was a major motivation for this work.

This work presents, in detail, a model for multiobjective routing in MPLS networks formulated within the framework developed by the authors in [7], assuming that there are two classes of services (and different types of traffic flows in each

class), namely QoS and BE services. The flows of QoS type (first priority flows), when accepted by the network, have a guaranteed QoS level, related to the required bandwidth, while BE traffic flows, which are treated in the model as second priority flows, are carried by the network in order to obtain the best possible QoS level for the current network routing solution. Another feature of the routing model is the use of alternative routing: when a first choice route assigned to a given micro-flow, belonging to a certain traffic flow (corresponding to a “traffic trunk”) is blocked, a second choice route may be attempted. An important feature of this model is the use of hierarchical optimization typically with two optimization levels, including fairness objectives: the first priority objective functions refer to the network level objectives of QoS type flows, namely the total expected revenue and the maximal value of the mean blocking of all types of QoS flows; the second priority objective functions refer to performance metrics for the different types of QoS services and the total expected revenue associated with the BE traffic flows. Another important feature of the model is the use of an approximate stochastic representation of the traffic flows in the network, based on the use of the concept of effective bandwidth for macro-flows and on a generalized Erlang model for estimating the blocking probabilities in the arcs, similar to the one used in [29, 26]. After describing in detail the routing model in Sect. 2, including the underlying traffic model, we will present in Sect. 3 the theoretical foundations of a specialized heuristic strategy for finding “good” compromise solutions to the very complex bi-level routing optimization problem. This theoretical foundation is based on a conjecture concerning the definition of marginal implied costs for QoS flows and BE flows, presented for the first time in this paper, which is an extension and adaptation of earlier definitions of implied cost for single service networks with alternative routing in [18]. The structure of the heuristic procedure for resolving the problem is analogous to the one described in detail in [7, 26]. The new version of the heuristic, presented here, is based on a constrained bi-objective shortest path model, the objective functions of which are QoS or BE marginal path implied costs, depending on the class of the routed traffic, and path blocking probabilities. Also, in Sect. 4, a description of a first version of this heuristic will be presented, and some preliminary results for a test network will be revealed in Sect. 5.

2 Description of the Routing Model

The present model can be considered as an application of the multiobjective modeling framework for MPLS networks described by the authors in [7]. This framework (or “meta-model”) uses hierarchical optimization with up to three optimization levels: the first priority objective functions refer to the global network level; the second priority objective functions refer to performance metrics for the different types of services supported by the network; the third priority functions are concerned with performance metrics for the micro-flows of packet streams of the same FEC.

It is a network-wide routing optimization approach of a new type, in the form of a hierarchical multiobjective optimization model, which takes into account the nature and relations between the adopted objective functions related to the different types of traffic flows associated with different services. We would like to note that various multiobjective models previously proposed use objective functions chosen to reflect only indirectly network technical-economic objectives. A typical example is the minimization of a utilization cost for all arcs expressed, through empirical functions, in terms of the occupied bandwidth as in [14, 13, 12, 20]. In fact, the pursued objective is to optimize the total traffic carried in the network or the associated expected revenue. One can say that this type of approach is just a rough approximation to the ‘hidden’ (or implicit) objective function the model seeks to reflect, especially taking into account the random nature of traffic patterns, even in stationary or quasi-stationary network working conditions. Instead, our model considers an explicit representation of the most relevant technical-economic objectives in a network-wide routing optimization, such as the total expected revenue (expressed in terms of the traffic carried of all service types). This aspect of the modeling approach is in line with the school of thought adopted by [18, 19, 29], in the context of single-objective routing models.

We propose a hierarchy of objective functions by considering in a first approach two levels of optimization with several objective functions in each level. The first level (first priority) includes objective functions formulated at network level for the QoS type traffic and considering the combined effect of all types of traffic flows in the network. The second level refers to average performance metrics of the QoS traffic flows associated with the different types of services supported by the network and the expected revenue of the BE traffic. An important feature of the model is the explicit consideration, as objective functions, of ‘fairness’ objectives, at the two levels of optimization. These are objectives of min-max type and seek to make the most of the proposed multiobjective formulation. In previous formulations of routing models for these networks, such aims related to fairness are usually not considered explicitly in any form or just represented through constraints on certain performance metrics. Another important feature of the model is the stochastic representation of the traffic flowing in the network as described in [7, 29].

We will consider two *classes* of services, namely QoS corresponding to services with certain guaranteed QoS levels, and BE, where the corresponding traffic flows are routed seeking to obtain the best possible quality of service but not at the cost of the QoS of the QoS traffic flows (first priority traffic flows). The *service types* in each class are grouped in the sets \mathcal{S}_Q (for QoS service types) and \mathcal{S}_B (for BE service types), and the traffic flows of each service type $s \in \mathcal{S}_Q$ or $s \in \mathcal{S}_B$ may differ in important attributes, namely the required bandwidth.

The consideration of two (or more) classes of traffic flows in a routing model is a complex issue and different heuristic approaches have been proposed in the literature. Examples of these approaches, typically flow-oriented models, are in [20, 23, 21]. As for network-wide optimization approaches, [30] describes a bi-objective routing model using lexicographic optimization where a primary objective function is the weighted sum of the carried bandwidth associated with QoS traffic flows and

a secondary objective function of the same type is defined for the BE traffic. A heuristic procedure based on a decomposition technique and multicommodity flow programming is developed for obtaining solutions to the problem.

Some definitions relevant to the model are now formally introduced.

Definition 1. A *traffic flow* is a mathematical entity specified by $f_s = (v_i, v_j, \bar{\gamma}_s, \bar{\eta}_s)$ for $s \in \mathcal{S} = \mathcal{S}_Q \cup \mathcal{S}_B$ that corresponds to a stochastic process, in general a marked point process, that describes the arrivals and basic requirements of μ -flows, originated at the MPLS ingress node v_i and destined to the MPLS egress node v_j , using the same LSP and characterized by the vectors of ‘attributes’ $\bar{\gamma}_s$ and $\bar{\eta}_s$ for service type s .

The vector $\bar{\gamma}_s$ describes the traffic engineering attributes of flows of service type s and the vector $\bar{\eta}_s$ enables the representation of the mechanism(s) of admission control to all links l_k in the network by calls of flow f_s . The set of all traffic flows of type s will be denoted by \mathcal{F}_s .

In the teletraffic modeling approach considered here, these attributes include the required effective bandwidth d_s and the mean duration $h(f_s)$ of a μ -flow in f_s . In our model a ‘ μ -flow’ corresponds to a ‘call’, the term call being used in its broadest sense, that is, as a connection request with certain features. The use of the concept of effective bandwidth [17] in the present context (MPLS networks with explicit routes) was earlier proposed in [29] and used in [25, 26]. This enables an upper level network representation in the traffic plane level, through an equivalent multirate loss traffic network.

Consider that we have an approximate teletraffic model that is capable of estimating the node-to-node blocking probabilities $B(f_s)$ for all flows f_s of all service types.

Definition 2. A *routing plan* \bar{R} is a set of loopless paths, for all network services and all flows, which corresponds to a possible routing pattern in the network, assuming that up to $M - 1$ alternative routes may be attempted by any connection request of f_s : $\bar{R} = \cup_{s=1}^{|\mathcal{S}|} R(s)$ for all the network services, where $|\mathcal{S}|$ is the total number of services, $R(s) = \cup_{f_s \in \mathcal{F}_s} R(f_s)$, $s \in \mathcal{S}_Q \cup \mathcal{S}_B$ and $R(f_s) = (r^p(f_s))$, $p = 1, \dots, M$.

In the present model (one-stage alternative routing), $M = 2$.

This means that for each flow f_s the first choice route $r^1(f_s)$ will be used unless it is blocked as a result of one of its links l_k not having the required available bandwidth d_s (or as prescribed by a general probabilistic availability function ψ_{ks}). If $r^1(f_s)$ is blocked then the second choice route $r^2(f_s)$ will be attempted by the connection request and the request will be blocked only if $r^2(f_s)$ is also blocked.

Let $A(f_s)$ represent the mean of the traffic offered by traffic flow f_s in Erl and A_s^o represent the total traffic offered by the flows of the service type s , $A_s^o = \sum_{f_s \in \mathcal{F}_s} A(f_s)$ [Erl].

Definition 3. The *average blocking probability for all traffic flows of type s* , for a given routing plan \bar{R} , is $B_{ms} = \frac{1}{A_s^o} \sum_{f_s \in \mathcal{F}_s} A(f_s) B(f_s)$.

In particular, $B_{ms|Q} = B_{ms}$ for given $s \in \mathcal{S}_Q$.

Definition 4. The *maximal blocking probability among all traffic flows f_s of QoS class and type s* , is $B_{Ms|Q} = \max_{f_s \in \mathcal{F}_s} \{B(f_s)\}, s \in \mathcal{S}_Q$.

Definition 5. The *maximal average blocking probability among all QoS service types*, is $B_{Mm|Q} = \max_{s \in \mathcal{S}_Q} \{B_{ms}\}$.

Let $w(f_s)$ denote the expected revenue associated with calls of a generic flow f_s and A_s^c be the total carried traffic by traffic flows of type s , $A_s^c = \sum_{f_s \in \mathcal{F}_s} A(f_s)(1 - B(f_s)) = A_s^o(1 - B_{ms})$ [Erl]. Further assume the usual simplification $w(f_s) = w_s, \forall f_s \in \mathcal{F}_s$.

Definition 6. The *total expected revenues associated with QoS(BE) traffic flows*, are given by $W_{Q(B)} = \sum_{s \in \mathcal{S}_{Q(B)}} A_s^c w_s$.

In the framework of the meta-model [7] we may formulate a two-level multiobjective routing optimization problem by separating the total expected revenue in two parts: W_Q for the traffic flows of QoS type and W_B for the traffic flows of BE type, as defined above, and by considering explicitly performance optimization of QoS service types. While W_Q will be a first priority objective function, together with the maximal blocking probability for all QoS service types, $B_{Mm|Q}$, W_B will be a second level objective function. This seeks to guarantee that the routing of BE traffic, in a quasi-stationary situation, will not be made at the cost of a decrease in revenue or of an increase in the blocking probability of QoS traffic flows.

The second level of optimization also concerns QoS service types and includes $2|\mathcal{S}_Q|$ objective functions to be minimized, the mean blocking probability $B_{ms|Q}$ for flows of type $s \in \mathcal{S}_Q$, and the maximal blocking probability $B_{Ms|Q}$, defined above. Note that $B_{Ms|Q}$ represents the fairness objective defined for each service type $s \in \mathcal{S}_Q$.

These considerations led to the following formulation of a *two-level hierarchical optimization problem for two service classes*:

Problem P-M2-S2

$$\begin{array}{l}
 \text{1st level} \left\{ \begin{array}{l} \text{QoS - Network objectives: } \min_{\bar{R}} \{-W_Q\} \\ \min_{\bar{R}} \{B_{Mm|Q}\} \end{array} \right. \\
 \text{2nd level} \left\{ \begin{array}{l} \text{QoS - Service objectives: } \min_{\bar{R}} \{B_{ms|Q}\} \\ \min_{\bar{R}} \{B_{Ms|Q}\} \\ \forall s \in \mathcal{S}_Q \\ \text{BE - Network objective: } \min_{\bar{R}} \{-W_B\} \end{array} \right. \\
 \text{subject to the equations of the underlying traffic model.}
 \end{array}$$

Here the acronym P-M2-S2 stands for ‘**P**roblem - **M**ultiobjective with **2** optimization hierarchical levels - with **2** Service classes’.

It is important to note that while QoS and BE traffic flows are treated separately in terms of objective functions in order to take into account their different priority in

the optimization model, the interactions among all traffic flows are fully represented in the model. This is guaranteed by the adopted traffic modeling approach underlying the optimization model. This is another major difference in comparison to other routing models proposed for networks with two service classes.

The definition and calculation of the parameters in the expressions are given in [8, Appendix A].

It should be noted that this model is a simplification of the general model for QoS and BE service classes outlined by the authors in [7, Sect.3.3]. In the addressed routing optimization model P-M2-S2 only the macro level representation was considered, having in mind the avoidance of increased complexity in a model, which is by itself very complex, resulting from the inclusion of a third optimization level in the routing model, as well as the corresponding additional computational burden.

The traffic modeling approach is the one used in [7] and earlier in [29, 26] for tackling the calculation of blocking probabilities experienced by the traffic flows in network links. It is based on the concept of *effective bandwidth* (see underlying theory in [17]), in association with the definition of MPLS explicit routes. The effective bandwidth can be considered as a stochastic measure of the utilization of network resources enabling the representation (in an approximate manner) of the effects of the variability of the rates of different traffic sources, as well as the effects of statistical multiplexing of different traffic flows in a network. This conceptual tool was used in routing optimization models of multiservice networks of various types as in [29, 26]. Hence, the network may be represented in the traffic plane by a multiclass loss traffic network, equivalent to a multirate traffic circuit-switched network.

The basic calculation sub-model enables the blocking probabilities B_{ks} , for connection requests of service type s in link l_k , to be obtained in the form $B_{ks} = \mathcal{L}_s(\bar{d}_k, \bar{\rho}_k, C_k)$. \mathcal{L}_s represents the basic function (implicit in the analytical model) that gives the marginal blocking probabilities, B_{ks} , in terms of $\bar{d}_k = (d_{k1}, \dots, d_{k|\mathcal{S}|})$ (vector of equivalent effective bandwidths), $\bar{\rho}_k = (\rho_{k1}, \dots, \rho_{k|\mathcal{S}|})$ (vector of reduced traffic loads ρ_{ks} offered by flows of type s to l_k) and the link capacity C_k .

This approximation was suggested in [29] in the context of off-line single-objective multiservice routing optimization models and was also used in the multi-objective dynamic alternative routing model [26]. The use of very efficient and robust stochastic approximations is absolutely critical in a routing optimization model of this type, for tractability reasons. The detailed description of the traffic model can be seen in [7, 8].

3 Foundations of the Resolution Approach

In the hierarchical multiobjective network-wide optimization routing problem P-M2-S2 we will consider that the routing principle uses alternative routing i.e. the decision variables are the network routing plans \bar{R} .

The hierarchical multiobjective alternative routing problem in question is highly ‘complex’ as a result of two major factors: the strong interdependencies among all

objective functions (via the $\{B(f_s)\}$) and the interdependencies between the objective function parameters and the (discrete) decision variables \bar{R} (network route plans). All these interdependencies are defined by the underlying traffic model.

Concerning overall complexity it can be said that the simplest, ‘degenerated’ single objective version of the problem, corresponding to the single objective function W_Q , one single service and no alternative routing ($M = 1$) is NP-complete in the strong sense, as shown in [11]. Note that our model is a bi-level, multiobjective formulation of this type of problem. This and the interdependencies among the objective functions are a strong indication of extreme intractability of the problem.

Concerning the possible conflict between the objective functions in P-M2-S2, it can be said that in many situations, the maximization of W_Q entails a deterioration on $B(f_s)$, $s \in \mathcal{S}_Q$, for “small” intensity traffic flows $A(f_s)$ which tends to increase $B_{Ms|Q}$ and, as a result, $B_{Mm|Q}$. In single-objective routing models this effect is usually tackled by imposing upper bounds on the values $B(f_s)$. These relations between objective functions of this type have been analyzed in [27]. Note that this is a major factor to justify the interest and potential benefit in using multiobjective approaches when dealing with this type of routing problem.

The resolution (in a multicriteria analysis sense) of the routing model P-M2-S2 will be performed by a heuristic approach, a first version of which is presented in the next section. This heuristic is the extension and adaptation to this problem of the heuristic procedure described in [9] and [26].

The heuristic developed for problem P-M2-S2 is based on the calculation of solutions of a bi-objective shortest path problem. In this problem the path metrics to be minimized will be the marginal implied costs (as defined according to the following analysis) and blocking probabilities.

The implied cost c_{ku} resulting from the acceptance of a call of flow f_u in link l_k is a powerful mathematical concept in routing optimization in circuit-switched networks, and was originally proposed by Kelly [18] for single-rate traffic networks. It was extended to single route (i.e. without alternative routing) multirate traffic networks in [15] and [29]. It can be defined as the expected value of the loss of revenue in all network traffic flows which may use link l_k resulting from the acceptance of a call from f_u associated with the decrease in the capacity of this link. Therefore the implied cost measures the knock-on effects on all network routes (of all traffic flows) resulting from the acceptance of a call from f_u in a link l_k . The authors have adapted in [9] and [26] the definition of c_{ku} to multirate loss networks with alternative routing by extending the corresponding expression given for single-service networks in [18]. This extension implies that the c_{ku} can be calculated from the equations:

$$c_{ku} = \sum_{s \in \mathcal{S}} \frac{1}{1 - B_{ks}} \zeta_{kus} \left[\sum_{f_s: l_k \in r^1(f_s)} \lambda_{r^1(f_s)} (s_{r^1(f_s)} + c_{ks}) + \sum_{f_s: l_k \in r^2(f_s)} \lambda_{r^2(f_s)} (s_{r^2(f_s)} + c_{ks}) \right] \quad (1)$$

with

$$\begin{aligned} s_{r^2(f_s)} &= w(f_s) - \sum_{l_j \in r^2(f_s)} c_{js} \\ s_{r^1(f_s)} &= w(f_s) - \sum_{l_j \in r^1(f_s)} c_{js} - (1 - L_{r^2(f_s)})s_{r^2(f_s)} \end{aligned} \quad (2)$$

where $s_{r^p(f_s)}$ is the surplus value of a call on route $r^p(f_s)$, $\lambda_{r^p(f_s)}$ is the marginal traffic carried on $r^p(f_s)$, $L_{r^p(f_s)}$ is the blocking probability for calls of f_s on route $r^p(f_s)$ ($p = 1; 2$), considering that $r^1(f_s)$ and $r^2(f_s)$ are disjoint paths, and $\zeta_{kus} = \mathcal{L}_s(\bar{d}_k, \bar{\rho}_k, C_k - d_{ku}) - \mathcal{L}_s(\bar{d}_k, \bar{\rho}_k, C_k)$ is the increase in the congestion for type s calls on link l_k originated by a decrease in the arc capacity because of the acceptance of a type u call.

The calculation of the implied costs in this form is based on the following conjecture, which is an extension to multirate loss networks with alternative routing of the results in [18, Sect.7] and [28, Sect.III].

Conjecture A: In multirate networks with (one-stage) alternative routing the sensitivity of the revenue W_T with respect to the traffic $A(f_s)$ being offered to a pair of routes (r^1, r^2) , when an approximation to the expected revenue is calculated from the solution of the fixed point equations in B_{ks} , can be written

$$\begin{aligned} \frac{\partial W_T}{\partial A(f_s)} &= (1 - L_{r^1(f_s)}) \left(w_{r^1(f_s)} - \sum_{l_k \in r^1(f_s)} c_{ks} \right) \\ &+ L_{r^1(f_s)}(1 - L_{r^2(f_s)}) \left(w_{r^2(f_s)} - \sum_{l_k \in r^2(f_s)} c_{ks} \right) \end{aligned} \quad (3)$$

where the c_{ks} are the implied costs and these satisfy the system of equations (1)-(2).

The validity of this conjecture implies that the implied cost c_{ku} can be interpreted as the exact mathematical measure of the ‘knock-on’ effects on all network routes of all network flows resulting from the acceptance of a call from f_u in link l_k , in the terms stated above. This in turn is consistent with the calculation of the c_{ks} through the system (1), by applying the theory developed in [18].

It will be further assumed that $w_{r^1(f_s)} = w_{r^2(f_s)} = w(f_s)$. The fixed point equations in this statement result from the traffic model and constitute a system of implicit non-linear equations enabling the calculation of the B_{ks} in terms of link capacities (expressed through matrix $\bar{C} = [C_k]$), the offered traffic matrix $\bar{A} = [A(f_s)]$, and the current network routing plan \bar{R} . The calculation of c_{ks} through (1)-(2) also implies the solution of a system of equations.

$$B_{ks} = \beta_{ks}(\bar{B}, \bar{C}, \bar{A}, \bar{R}) \quad \text{and} \quad c_{ks} = \alpha_{ks}(\bar{c}, \bar{B}, \bar{C}, \bar{A}, \bar{R}) \quad (4)$$

with $k = 1, \dots, |\mathcal{L}|$; $s = 1, \dots, |\mathcal{S}|$, $\bar{B} = [B_{ks}]$ and $\bar{c} = [c_{ks}]$. The numerical resolution of the systems (4) is performed by fixed point iterators, for given \bar{C}, \bar{A} and \bar{R} .

In [9] and [26], the authors formulated a bi-level *multiple objective dynamic alternative routing problem for multiservice networks* for a single service class with multiple service types:

Problem $\mathcal{P}_G - S$

$$\left\{ \begin{array}{l} \text{Network level: } \min_{\bar{R}} \{-W_T\} \\ \min_{\bar{R}} \{B_{Mm} = \max_{s \in \mathcal{S}} \{B_{ms}\}\} \\ \text{Service level: } \min_{\bar{R}(s)} \{B_{ms}\}, s = 1, \dots, |\mathcal{S}| \\ \min_{\bar{R}(s)} \{B_{Ms} = \max_{f_s \in \mathcal{F}_s} \{B(f_s)\}\}, s = 1, \dots, |\mathcal{S}| \end{array} \right.$$

subject to the equations of the teletraffic model enabling the calculation of $\{B(f_s)\}$ in terms of $\{A(f_s)\}$ and \bar{R} , where W_T is the total expected network revenue associated with the traffic carried by all service types, B_{ms} is the average blocking probability for all traffic flows of service type s , B_{Ms} is the maximal value of those blocking probabilities, and B_{Mm} is the maximal value among the B_{ms} and is, together with W_T , the first priority objective function. This routing model can be considered as an application of the meta-model associated with P-M3-S2 (**Problem - Multiobjective with 3 optimization hierarchical levels - with 2 Service classes**) [7] and as a particular case of the addressed model, P-M2-S2, by considering only one service class.

The resolution approach to $\mathcal{P}_G - S$ was based on the calculation of solutions to the bi-objective shortest path problem $\mathcal{P}^{(2)}$, formulated for every end-to-end flow f_s , $\min_{r(f_s) \in \mathcal{D}(f_s)} \{m^n(r(f_s)) = \sum_{k \in r(f_s)} m_{ks}^n\}_{n=1,2}$, where $m_{ks}^1 = c_{ks}$ and $m_{ks}^2 = -\log(1 - B_{ks})$, and $\mathcal{D}(f_s)$ is the set of feasible loopless paths for flow f_s , resulting from traffic engineering constraints. The logarithmic function is used to transform the blocking probability into an additive metric.

The use of this constrained bi-objective shortest path problem as a basis for the resolution approach to the network problem $\mathcal{P}_G - S$ relies on the fact that the metric blocking probability tends (at a network level) to minimize the maximal node-to-node blocking probabilities $B(f_s)$ while the metric implied cost tends to maximize the total average revenue W_T (see [10] and [27]). When one states that using the minimization of path implied cost ‘tends’ to maximize W_T this would be in rigor only valid if the choice of such an ‘optimal’ path for a given f_s would not change in any form the network working condition, an assumption that is not true, having in mind the interdependencies among $\{c_{ks}\}$, $\{B_{ks}\}$ and \bar{R} (see (4)). This is the ultimate source of difficulty in devising a heuristic based on this principle, as outlined in the next section.

Nevertheless it was possible to develop a heuristic approach based on this principle that gave very good results when compared with reference routing methods like RTNR (Real Time Network Routing) of AT&T, and DAR (Dynamic Alternate Routing), aimed at maximizing the total expected revenue, as shown in [26].

In order to extend this resolution principle to the problem P-M2-S2 we need to extend the definition of implied costs to a network with two service classes. For this purpose we propose the following definition of *marginal implied costs* associated

with QoS (BE) traffic, by extending the original interpretation of implied costs by Kelly [18] to a multirate loss network with two service classes. Hence we will define the *marginal implied cost for QoS traffic*, c_{ku}^Q , associated with the acceptance of a connection (or “call”) of traffic f_u of any service type $u \in \mathcal{S}$ on a link l_k , as the expected value of the traffic loss induced on all QoS traffic flows resulting from the capacity decrease in link l_k . In an analogous form one can define the marginal implied cost c_{ku}^B for BE traffic associated with the acceptance of a connection of traffic flow f_u on link l_k .

We will assume, as a conjecture, that the marginal implied costs for QoS (BE) traffic can be estimated by solving a system of equations analogous to (1)-(2), by restraining the summation on the right hand side of (1) to the service types of QoS (BE) class, respectively, by introducing the marginal surplus values $s_{r^i(f_s)}^{Q(B)}$ and the marginal revenues,

$$w^{Q(B)}(f_s) = \alpha^{Q(B)} w(f_s) \text{ with } \alpha^Q + \alpha^B = 1.0 \quad (5)$$

where the coefficients $\alpha^{Q(B)} \in]0.0; 1.0[$ satisfy the above normalization condition. This condition and the calculation of the marginal costs through those equations are consistent with the definition of the sensitivity of the marginal revenues associated with QoS and BE traffic, through expressions analogous to (3), as described in the following conjecture.

Conjecture B: In multirate networks with (one-stage) alternative routing and two service classes, the sensitivity of the revenues $W_{Q(B)}$ with respect to the traffic $A(f_s)$ being offered to a pair of routes (r^1, r^2) , when an approximation to the expected revenue is calculated from the solution of the fixed point equations in B_{ks} , can be written

$$\begin{aligned} \frac{\partial W_{Q(B)}}{\partial A(f_s)} = & (1 - L_{r^1(f_s)}) \left(w^{Q(B)}(f_s) - \sum_{l_k \in r^1(f_s)} c_{ks}^{Q(B)} \right) \\ & + L_{r^1(f_s)} (1 - L_{r^2(f_s)}) \left(w^{Q(B)}(f_s) - \sum_{l_k \in r^2(f_s)} c_{ks}^{Q(B)} \right) \end{aligned} \quad (6)$$

In fact, taking into account that $W_T = W_Q + W_B$, (6) and (5), together with the condition $c_{ks} = c_{ks}^Q + c_{ks}^B$, imply (from the additivity property of the derivatives) equation (3) (in conjecture A). The marginal expected revenues per call of f_s , $w^{Q(B)}(f_s)$ (such that $w^Q(f_s) + w^B(f_s) = w(f_s)$) in equation (6) may be interpreted as the part of the expected revenue $w(f_s)$ generated by a connection of f_s that is accepted by the network (for a given choice of the pair of routes $(r^1(f_s), r^2(f_s))$), that is assigned to the calculation of the sensitivity of the revenue from the point of view of traffic losses induced either in the QoS traffic flows or in the BE flows.

In the present model we will consider, as a first approach, $\alpha^Q = \alpha^B = 0.5$ so that no bias is induced in the calculation of the marginal costs through the choice of these factors.

This conjecture plays a central role in the theoretical foundation of the resolution approach, since its validity implies that the marginal implied costs associated with QoS(BE) traffic can be interpreted exactly as stated in the paragraph above. This also implies the consistency of the calculation of the $\{c_{ks}^{Q(B)}\}$ through a system analogous to (1), with the changes associated with the introduction of marginal coefficients $w^{Q(B)}(f_s)$.

The auxiliary bi-objective shortest path problem to be solved will have two possible configurations. The first one will use as link cost coefficients $m_{ks}^1 = c_{ks}^Q$ when one intends to obtain candidate solutions to improve the revenue of the QoS traffic, and the second one will use $m_{ks}^1 = c_{ks}^B$ when one seeks to improve the revenue associated with the BE traffic.

To solve this constrained shortest path problem we will use an adaptation of the previously developed algorithmic approach MMRA-S or Modified Multiobjective Routing Algorithm for multiservice networks, described in [9] and [26].

4 A First Heuristic Approach

Next we describe the main features of a heuristic procedure for solving the model in the context of application to the MPLS network used in [30]. The basic architecture of the heuristic is analogous to the MODR-S (Multiobjective Dynamic Routing for Multiservice networks) heuristic described in [9] and applied in [26], with some relevant adaptations. These adaptations and changes in the heuristic have to do, on the one hand, with the different type of the considered network topology, as shown in Figure 1, that unlike the one for which MODR-S was specified (a fully meshed network) has low connectivity. On the other hand, there are important differences in the objective functions as a result of the existence of two traffic classes, as previously analyzed.

The ‘core’ of the heuristic is the generation of candidate solutions $(r^1(f_s), r^2(f_s))$ for each f_s , where $r^1(f_s)$ is defined according to the rules described hereafter and $r^2(f_s)$ is typically obtained through a constrained bi-objective shortest path algorithm, devised for problem $\mathcal{P}^{(2)}$, MMRA-S2.

Having in mind the network topology and the need to make a further distinction between real-time QoS services (video and voice services) and non-real time QoS services (such as ‘premium data’ service), special rules were defined for the selection of candidate first choice routes $r^1(f_s)$. An important parameter defined from these rules is the maximal number of arcs D_s per route for each service type s .

In general, for QoS services, $r^1(f_s)$ is chosen as the direct arc whenever it exists. If no direct path exists, then for real-time QoS services one of the feasible paths with the least number of arcs is chosen. If there is more than one of these paths, the choice is made according to MMRA-S2, by using priority regions defined in

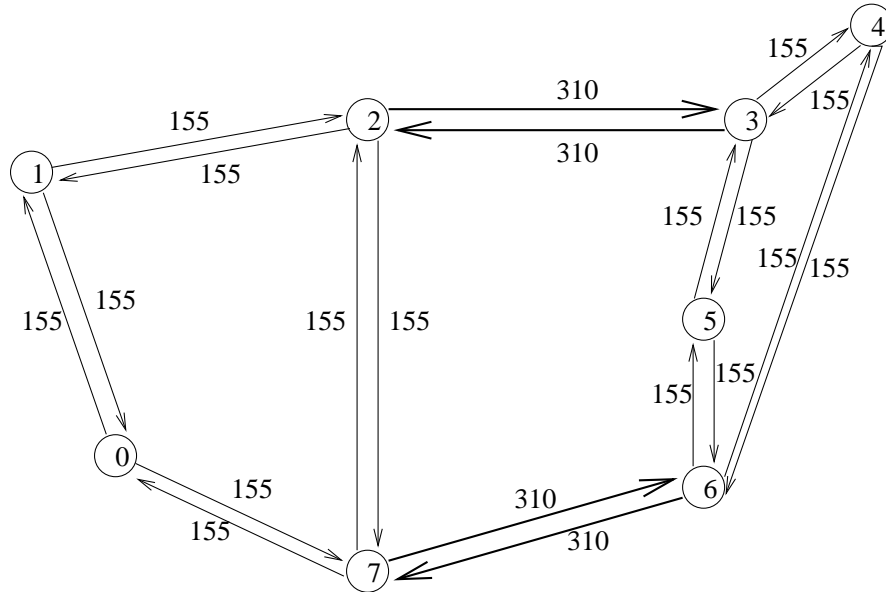


Fig. 1 Test network M1, proposed in [30], with the indication of the bandwidth of each arc C'_k , in Mbps

the objective function space of $\mathcal{P}^{(2)}$ (see [9, 26]). These criteria result from the more stringent constraints on delay and jitter of this type of services, and also having in mind an increase in reliability of the connections. For the remaining QoS services the initial choice of $r^1(f_s)$ is made by using the algorithm MMRA-S2 and the mentioned priority regions. A similar procedure is applied for obtaining $r^1(f_s)$ for BE traffic flows. Note that for BE traffic flows, the direct path, whenever it exists, is treated exactly as any other path, i.e. no preference is given to the direct path over the other paths. Concerning the calculation of the second choice routes $r^2(f_s)$ for QoS or BE traffic, this is made according to the MMRA-S2 algorithm. These alternative routes may be eliminated through a mechanism designated as Alternative Path Removal (APR) proposed in [26], in order to prevent performance degradation in overload conditions. Therefore, $r_s = r^2(f_s)$ is eliminated whenever $m^1(r_s) > \alpha^{Q(B)} \cdot d_s \cdot z_{APR}$ and $m^2(r_s) > -\log(0.7) \cdot z_{APR}$, where $z_{APR} \in [0.0; 1.0]$ is an empirical parameter which has to be adequately chosen through extensive experimentation with the model. The complete set of rules used in the given network to define candidate routes $(r^1(f_s), r^2(f_s))$ is described in [8, Appendix C].

As for the ‘core’ algorithm MMRA-S2 it is basically an adaptation to the present model of the bi-objective constrained shortest path algorithm in [26], which is an extension of the algorithm in [10] to a multiservice environment. The main features of MMRA-S are now briefly reviewed. Note that in general there is no feasible solution which minimizes both objective functions of $\mathcal{P}^{(2)}$ simultaneously. Since there is no guarantee of the feasibility of this ideal optimal solution, the resolution

of this routing problem aims at finding a best compromise path from the set of non-dominated solutions, according to some relevant criteria previously defined. In this context, since path computation and selection have to be fully automated, such criteria are embedded in the working of the algorithm MMRA [10, 27] via preference regions in the objective function space.

The purpose of the MMRA version used in the present context is to calculate solutions to problem $\mathcal{P}^{(2)}$, and is a variant of the algorithm proposed in [1] for a bi-objective shortest path problem of the same type, adapted to the requirements and specifics of the multiobjective dynamic routing model MODR-S [26]. The approach given in [1] (inspired by the one presented in [32] and [6] concerning a procedure to search interactively non-dominated paths) enables the calculation and selection of non-dominated paths in the framework of a routing control mechanism. The procedure satisfies this requirement by integrating the K -shortest paths algorithm [24] and a special concept designated as “soft constraints” (that is constraints not directly incorporated into the mathematical model). The main features of this approach are: i) the representation of QoS requirements through soft constraints corresponding to *requested* and *acceptable thresholds* for each QoS metric; ii) the consideration of this type of thresholds defines priority regions in the objective function space in which non-dominated solutions are searched for; iii) the non-dominated paths are computed by means of an extremely efficient K -shortest path algorithm proposed in [24], designated as MPS algorithm; iv) the adaptation of the preference thresholds to time-varying network working conditions, as required in dynamic routing applications. Further details on the MODR-S teletraffic model and MMRA, in a multiservice network, can be seen in the report [25].

This algorithm was adapted straightforwardly to the present routing model, by incorporating in the definition of the feasible route set $\mathcal{D}(f_s)$ and in the route selection procedure the rules described in detail in [8, Appendix C]. Also the coefficients of the second objective function are the marginal implied costs, either c_{ks}^Q or c_{ks}^B depending on the expected revenue (associated with QoS or BE traffic) we are seeking to improve.

Note that successive application of MMRA-S2 to every traffic flow does not lead to an effective (even less robust) resolution approach to the network routing problem P-M2-S2. The essential reason for this is an instability phenomenon that arises in such a path selection procedure, expressed by the fact that the route sets \bar{R} tend to oscillate between certain solutions, some of which may lead to poor global network performance under the prescribed criteria. This is associated with the complexity and interdependencies in the network problem P-M2-S2, namely the interdependencies between $\{c_{ks}^{Q(B)}\}$ and $\{B_{ks}\}$ and between these two sets of parameters and the current total route set \bar{R} .

A general idea of the heuristic is to seek, for each service, a routing solution $\bar{R}(s)$ which may lead to a better performance in terms of W_B , $B_{ms|Q}$ and $B_{Ms|Q}$, $s \in \mathcal{S}_Q$, hence leaving network resources available for traffic flows of other services so that the solutions selected at each step may enable an improvement on the higher level objective functions W_Q and $B_{Mm|Q}$. Therefore the heuristic is constructed in order to seek, firstly for each QoS service and beginning with the higher bandwidth services

(considering $s = 1, \dots, |\mathcal{S}_Q|$) and, secondly, for each BE service and beginning with the higher bandwidth services ($s = |\mathcal{S}_Q| + 1, \dots, |\mathcal{S}|$), solutions which dominate the initial one, in terms of $B_{ms|Q}$ and $B_{Ms|Q}$ for QoS services and in terms of W_B for BE services, while not worsening any of the network main metrics, W_Q and $B_{Mm|Q}$ (taking into account the optimization priorities in P-M2-S2).

The basis of the heuristic approach (similarly to MODR-S [26]) is to search for the subset of the path set $\bar{R}^a = \cup_{s=1}^{|\mathcal{S}|} \bar{R}^a(s) : \bar{R}^a(s) = \{(r^1(f_s), r^2(f_s)), f_s \in \mathcal{F}_s\}$, the elements of which should be possibly changed in the next route improvement cycle. Detailed analysis and extensive experimentation with MODR-S led us to propose [27, 26] a specific criterion for choosing candidate paths for possible routing improvement by increasing order of a function $\xi(f_s)$ of the current $(r^1(f_s), r^2(f_s))$. The criterion depends explicitly on the first choice path $r^1(f_s)$ and on the alternative path $r^2(f_s)$. The adaptation of this criterion to the present model P-M2-S2 considers:

$$\xi(f_s) = F_C^{Q(B)}(f_s) F_L(f_s) \quad (7)$$

if the effect over QoS(BE) traffic is being considered, and where $F_C^{Q(B)}(f_s) = (n_2 - n_1) c_1^{Q(B)} + c_{r^1(f_s)}^{Q(B)} - c_{r^2(f_s)}^{Q(B)}$, $F_L(f_s) = 1 - L_{r^1(f_s)} L_{r^2(f_s)}$, $c_{r(f_s)}^{Q(B)} = \sum_{k \in r(f_s)} c_{ks}^{Q(B)}$ and $c_1^{Q(B)} = \frac{1}{n_1} \sum_{k \in r^1(f_s)} c_{ks}^{Q(B)} = \frac{1}{n_1} c_{r^1(f_s)}^{Q(B)}$.

The aim of the factor $F_C^{Q(B)}(f_s)$ is to favor (concerning the interest in changing the second choice route when seeking to improve W_Q or W_B) the flows for which the second choice route has a high implied cost and the first choice route a low implied cost. The factor $(n_2 - n_1)$ was introduced for normalizing reasons having in mind that $r^1(f_s)$ has n_1 arcs and $r^2(f_s)$ n_2 arcs, in the considered network. The aim of the second factor $F_L(f_s)$ is to favor the choice of the flows with worse current blocking probability. In the cases where overload conditions led to the elimination of the alternative path (see explanation above), $F_C^{Q(B)}(f_s) = c_{r^1(f_s)}^{Q(B)}$ and $F_L(f_s) = 1 - L_{r^1(f_s)}$.

A second point to be tackled by the heuristic procedure is to specify how many (and which) of the routes with smaller values of $\xi(f_s)$ should possibly be changed by applying MMRA-S2 once again. For this purpose the effect of each candidate route, in terms of the relevant objective functions, is anticipated by solving the corresponding analytical model.

The heuristic is initialized with a set of paths $(r^1(f_s), r^2(f_s))$ that was defined without any previous optimization. The quality of the final solution obtained with the heuristic is dependent on the quality of the initial one. The initial routing plan \bar{R}_o to be used in the heuristic considers that the $r^1(f_s)$ are chosen as follows: the initial path chosen for every flow f_s is the shortest one (that is, the one with minimum number of arcs); if there is more than one shortest path, we choose the one with maximal bottleneck bandwidth (i.e. the minimal capacity of its arcs); if there is more than one shortest path with equal bottleneck bandwidth, the choice is arbitrary. The alternative routes $r^2(f_s)$ are chosen from the candidate paths obtained from MMRA-S2 according to the procedures of solution analysis and selection explained above and in the next paragraphs.

In order to have a “good” initial solution some alternative paths must be eliminated from the initial set of paths. According to the criterion of elimination of the alternative paths proposed in [9] and [26], all paths $r^2(f_s)$ satisfying

$$B(f_s) > \frac{\sum_{f_s \in \mathcal{F}_s} B(f_s)}{|\mathcal{F}_s|} \text{ or } B(f_s) > 10\% \quad (8)$$

should be eliminated ($|\mathcal{F}_s|$ is the number of flows of type s). This procedure keeps only ‘good’ alternative paths in the initial solution, and it seeks to improve both the service performances (especially for the services with higher bandwidth demands) and the network main performance metrics. Note that if the final solution of the heuristic does not dominate the initial one (before the elimination of some alternative paths) in terms of the first level objective functions, then it is this initial solution that should be adopted. However, this situation never occurred in all the tests performed.

The heuristic starts off with a cycle that covers all the services, beginning with the QoS services $s = 1, \dots, |\mathcal{S}_Q|$ and ending with the BE services $s = |\mathcal{S}_Q| + 1, \dots, |\mathcal{S}|$. Note that QoS services are treated in the model as first priority traffic, and BE services as second priority traffic. Within each class of service, the algorithm begins with the types of services with higher bandwidth demands. Experience has shown that this ordering of the services generally leads to a better performance of the heuristic. Therefore, the heuristic is set up to find, for each service, and starting with the most demanding services, solutions that dominate the previous one with respect to the first level objective functions W_Q and $B_{Mm|Q}$, if possible without worsening the partial criteria for each service, $B_{ms|Q}$ and $B_{Ms|Q}$ for QoS services and W_B for BE services.

The two main cycles of the heuristic are improvement cycles of the objective functions. Two variables, $nPaths$ and $mPaths$, define the current number of paths, ordered according to $\xi(f_s)$, which are candidates for possible improvements in these two cycles. A specific service protection scheme to prevent excessive network blocking degradation in overload situations, the APR, is used, as described earlier. The parameter z_{APR} varies between 0.0 and 1.0, and its value is defined in the inner cycle of the heuristic.

Concerning the numerical complexity of this heuristic, it can be said that the instructions in the inner cycle of the procedure are executed at least $C_N = |\mathcal{S}|(6|\mathcal{F}_N| - 1)$, where $|\mathcal{F}_N| = \frac{1}{|\mathcal{S}|} \sum_{s \in \mathcal{S}} |\mathcal{F}_s|$ is the average number of traffic flows per service. This figure C_N is an indication of the heuristic numerical complexity just at the level of the ‘optimization’ procedures. One should note that each calculation of the objective functions and marginal costs (which are used as coefficients in the auxiliary bi-objective shortest path model) involves the numerical resolution of a large system of non-linear equations in $\{B_{ks}\}$ and $\{c_{ks}^{Q(B)}\}$ and such calculations have to be repeated whenever a candidate pair of paths $(r^1(f_s), r^2(f_s))$ is recalculated and analyzed, in terms of its impact on the objective function values.

This heuristic is formalized in the Appendix. A more detailed explanation is in the report [8].

5 Application of the Model

In this section, computational results obtained with the MOR-S2 heuristic in a network case study analogous to the one in [30], are presented. The “quality” of these results concerning W_Q was compared to results obtained with another model proposed in [30] for MPLS networks with two service classes that uses a lexicographic optimization formulation based on a deterministic MCF (*Multicommodity Flow*) model, which gives an upper bound to our objective function W_Q in P-M2-S2. For this purpose the network case study for two service classes in MPLS addressed in [30] was considered.

5.1 Application of the Model to a Network Case Study

In [30], a model is proposed for traffic routing and admission control in multiservice, multipriority networks supporting traffic with different QoS requirements. Having in mind a better understanding of the application case study we will begin with an overview of the relevant features of the model proposed in [30]. Instead of using stochastic traffic models in the calculation of paths, deterministic models are used, in particular mathematical programming models based on MCFs. These models are only a rough approximation in this context, and in fact they tend to under-evaluate the blocking probabilities. As a result, an adaptation of the original model was introduced in [30] in order to obtain more ‘correct’ models, that is models which give a better approximation in a stochastic traffic environment. The authors of [30] propose a simple technique to adapt the MCF model to a stochastic environment: the requested values of the flows bandwidths in the MCF model are compensated with a factor $\alpha \geq 0.0$, so as to model the effect of the random fluctuations of the traffic that are typical of stochastic traffic models. The higher the variability of the point processes of the stochastic model, the higher is the need for compensation and therefore the higher should α be. A deterministic traffic routing model based on MCFs is specified, where traffic splitting is used. This means that the required bandwidth of each flow may be divided by multiple paths from source to destination, allowing for a better load balancing in the network.

The objective functions of this problem to be maximized are the revenues W_Q and W_B , associated with QoS and BE flows. A bi-criteria lexicographic optimization formulation is considered, concerning the revenues W_Q and W_B , so that the improvements in W_B are to be found under the constraint that the optimal value of W_Q is maintained.

In the deterministic flow-based model [30], a base matrix $T = [T_{ij}]$ with offered bandwidth values from node i to node j [Mbps] is given. A multiplier $m_s \in [0.0; 1.0]$ with $\sum_{s \in \mathcal{S}} m_s = 1.0$ is applied to these matrix values to obtain the offered bandwidth of each flow $f_s = (v_i, v_j, \bar{\gamma}_s, \bar{\eta}(s, \mathcal{L}))$ to the network, $T(f_s) = m_s T_{ij}$. The transformation of this type of matrix into a matrix of traffic intensities, used in our stochastic traffic model, can be carried out as described in Appendix E.1 of the report [8]. The

model for calculating the blocking probabilities is described in Appendix B of the same report.

In the application example in [30], results for the QoS flows revenue W_Q are presented for three values of α : $\alpha = 0.0$ corresponds to the deterministic approach; $\alpha = 0.5$ is the compensation factor when calls arrive according to a Poisson process, service times follow an exponential distribution and the network is critically loaded; and $\alpha = 1.0$ for traffic flows with higher ‘variability’. The results for the BE revenue W_B are not presented, but for $\alpha = 0.0$ the maximum value is 79.33% of the maximum possible value W_B^{\max} , because 20.67% of the traffic is not even admitted to the network due to an admission control scheme. The results for the revenues obtained from the information provided in [30] are $W_{Q|\alpha=0.0}^{[30]} = 65156.00$ and $W_{B|\alpha=0.0}^{[30]} \leq 17462.50$; $W_{Q|\alpha=0.5}^{[30]} = 60829.72$; $W_{Q|\alpha=1.0}^{[30]} = 56338.65$.

Note that the revenue values W_Q in the model [30] should be viewed as upper bounds on the QoS revenue values of the problem P-M2-S2, because of the differences between the two optimization problems and the important differences in the underlying routing control and traffic models, previously referred to. In fact an important feature of the resolution approach of the routing problem in [30] is the admission control of BE traffic flows at the first stage of resolution, so the BE traffic that is actually offered to the network is the fraction of traffic that was not rejected by the admission control. Also note the absence of alternative routing as well as the use of traffic splitting. Therefore, for a specific traffic matrix, the model in [30] tends to obtain smaller values of blocking probability by comparison with a situation without admission control, and this tends to favor higher global revenues. Also the traffic representation, even in the approximated stochastic model [30, Sect.5.4], is a bit rough and tends to under-evaluate the blocking probabilities and to over-estimate the revenues.

5.2 Some Experimental Results

The test network M1 proposed in [30] is displayed in Figure 1. It has $N = 8$ nodes, with 10 pairs of nodes linked by a direct arc. The network has a total of $|\mathcal{L}| = 20$ unidirectional arcs, one in each direction for every pair of adjacent nodes. The bandwidth of each arc C'_k is shown in Figure 1. The number of channels C_k (with basic capacity $u_0 = 16$ kbps) is $C_k = \left\lceil \frac{C'_k}{u_0} \right\rceil$.

There are $|\mathcal{S}| = 4$ service types with the features described in Table 1. The value of $d_s = \frac{d'_s}{u_0}$ [channels] $\forall s \in \mathcal{S}$ presented in the table (where d'_s is the required bandwidth in kbps) is calculated with $u_0 = 16$ kbps. Note that $w_s = d_s, \forall s \in \mathcal{S}$.

The traffic flow data information provided by [30] is a base matrix $T = [T_{ij}]$ with offered bandwidth values [Mbps] and a multiplier applied to these matrix values to obtain the offered bandwidth of each flow. Given this information and the variability compensation equations, the values of $A(f_s)$, the average number of offered μ -flows

Table 1 Service features on the test network M1

Service	Class	d'_s [kbps]	d_s [channels]	w_s	h_s [s]	D_s [arcs]	m_s
1 - video	QoS	640	40	40	600	3	0.1
2 - Premium data	QoS	384	24	24	300	4	0.25
3 - voice	QoS	16	1	1	60	3	0.4
4 - data	BE	384	24	24	300	7	0.25

of f_s , during the average service time of a μ -flow can be calculated as shown in [8, Appendix E.1].

In MOR-S2, an initial solution has to be chosen and applied as input data to the heuristic. We chose to consider an initial solution with only one path for each flow, i.e. without a second choice path as we concluded that this is more adequate, and leave it up to the heuristic to find an adequate solution with second choice paths. The initial solution is the same for all the services and the paths are symmetrical.

The objective function values for the initial and final solutions obtained with the MOR-S2 heuristic are in Table 2. The revenue values have 2 decimal places and the blocking probability values have 3 significant figures. The value of the QoS revenue in the final solution is also presented as a percentage of the optimal value obtained in [30].

Table 2 Objective function values for the initial and final solution for MOR-S2 on the test network M1

Objective Functions	$\alpha = 0.0$		$\alpha = 0.5$		$\alpha = 1.0$	
	Initial	Final	Initial	Final	Initial	Final
W_Q	54803.69	64330.56*	51785.21	60097.78†	49010.41	55978.80‡
$B_{Mm Q}$	0.413	0.135	0.413	0.0962	0.405	0.0582
$B_{m1 Q}$	0.413	0.135	0.413	0.0962	0.405	0.0582
$B_{m2 Q}$	0.314	0.0159	0.296	0.00811	0.275	0.00279
$B_{m3 Q}$	0.0198	0.00489	0.0174	0.00263	0.0150	0.000436
$B_{M1 Q}$	0.912	0.848	0.882	0.628	0.841	0.440
$B_{M2 Q}$	0.766	0.0427	0.722	0.0305	0.667	0.0111
$B_{M3 Q}$	0.0585	0.0456	0.0517	0.0241	0.0446	0.0143
W_B	15106.57	17391.44	13787.49	17031.62	12445.64	16509.86

*)98.73% of $W_{Q|\alpha=0.0}^{[30]}$; †)98.80% of $W_{Q|\alpha=0.5}^{[30]}$; ‡)99.36% of $W_{Q|\alpha=1.0}^{[30]}$

The MOR-S2 heuristic manages to start off with an initial solution with poor values for the objective functions and still finishes with a solution with significantly better values. The values for all the objective functions, for all values of α , are improved through the heuristic. The QoS revenue of the final solutions are slightly worse than those of the optimal solution in [30], as expected. However, these QoS revenues can be considered very good as they stand for approximately 99% of the optimal values in [30]. Therefore, we can consider that MOR-S2 has managed to

find an adequate “good” compromise routing solution to the routing problem P-M2-S2. In fact these experimental results for three traffic matrices showed that the expected QoS revenue obtained with our heuristic is never less than 98.7% of that upper bound while a substantial improvement on the other objective functions could be obtained with respect to the initial solution, using only shortest path first choice routing, typical of Internet routing conventional algorithms.

6 Conclusions and Further Work

In the emergent MPLS technology for the Internet the implementation of connection-oriented services from origin to destination is possible. This feature in association with other functional capabilities of MPLS enables the implementation of advanced QoS-based routing mechanisms, namely by establishing “explicit routes” (determined at the originating node) for each traffic flow.

Having in mind these features and capabilities of MPLS routing it is possible to explore the multicriteria nature of the routing environment and associated metrics (of technical and economic nature), and devise multicriteria routing models capable of explicitly incorporating various network revenues and performance metrics, including fairness QoS objectives at the level of the services. This enables the formulation of multiobjective network-wide optimization routing models, namely hierarchical multicriteria models, for possible application at the top level of this type of networks.

In this work we described a bi-level multiobjective routing model in MPLS networks, formulated within the general modeling framework developed by the authors in [7], assuming that there are two classes of services (and different types of traffic flows in each class), namely QoS and BE services. The routing model also considers the possibility of using alternative routing when that is advantageous to the first priority objective functions. An important feature of this model is the use of hierarchical optimization with two optimization levels, including fairness objectives. Another feature of the model is the use of an approximate stochastic representation of the traffic flows in the network, based on the use of the concept of effective bandwidth for macro-flows and on a generalized Erlang model for estimating the blocking probabilities in the arcs. Also note that while QoS and BE traffic flows are treated separately in terms of objective functions in order to take into account their different priority in the optimization model, the interactions among all traffic flows are fully represented through the traffic model. This is another advantage in comparison to other routing models proposed for networks with two service classes.

We have also presented the theoretical foundations of a specialized heuristic strategy for finding “good” compromise solutions to the very complex bi-level routing optimization problem. This theoretical foundation was based on a conjecture concerning the definition of marginal implied costs for QoS flows and BE flows, which is an extension and adaptation of earlier definitions of implied costs for single-service networks with alternative routing in [18]. The structure of a first version

of a heuristic procedure for resolving the problem was described. The model was applied to a test network previously used in a benchmarking study [30] that uses a lexicographic optimization routing approach, including admission control for BE traffic, based on a deterministic traffic representation, the results of which can be considered as upper bounds with respect to the QoS traffic revenue. These preliminary results seem quite encouraging concerning the potential performance of a multicriteria routing model of this nature.

The major limitation of this type of model is its inherent great complexity and the associated computational burden, which constitute the reverse of its ‘ambitious’ features, namely, network-wide optimization, multiobjective nature with a significant number of objective functions, use of alternative routing and a stochastic representation of traffic flows of multiple service types. This makes, at present, its potential practical application restrained to networks with a limited number of nodes, such as the core and intermediate (metro-core) level networks of low dimension. The model could also be used as the basis of a periodic type dynamic routing method, similarly to MODR-S [26].

Further work on this model will involve the search for improvements in the heuristic procedure, through sensitivity analysis of the present version, or the possible development of metaheuristics for this very complex network routing problem. Finally a discrete event simulation platform will be developed, which will enable a more exact evaluation of the results of the heuristic in a stochastic dynamic environment closer to real network working conditions.

Appendix – Formalization of the MOR-S2 Heuristic

- I. $\bar{R}_a \leftarrow \bar{R}_o$
- II. Compute \bar{B} and $W_Q, B_{Mm|Q}$ for \bar{R}_a
- III. $W_Q^o \leftarrow W_Q, B_{Mm|Q}^o \leftarrow B_{Mm|Q}$
- IV. Eliminate the paths $r^2(f_s)$ in \bar{R}_a that verify (8)
- V. $\bar{R}_e \leftarrow \bar{R}_a$
- VI. Compute \bar{B} and $W_Q, B_{Mm|Q}$ for \bar{R}_a
- VII. $\max\{W_Q\} \leftarrow W_Q, \min\{B_{Mm|Q}\} \leftarrow B_{Mm|Q}$
- VIII. For $s = 1$ to $s = |\mathcal{S}|$
 1. $\bar{R}_a(s) \leftarrow \bar{R}_e(s); \bar{R}_*(s) \leftarrow \bar{R}_e(s)$
 2. Compute \bar{B} and $B_{ms}, B_{Ms}, s \in \mathcal{S}_Q$ or $W_B, s \in \mathcal{S}_B$ for \bar{R}_a
 3. $\min\{B_{ms}\}_{ini} \leftarrow B_{ms}, \min\{B_{Ms}\}_{ini} \leftarrow B_{Ms}, s \in \mathcal{S}_Q$ or $\max\{W_B\}_{ini} \leftarrow W_B, s \in \mathcal{S}_B$
 4. $mPaths \leftarrow |\mathcal{F}_s|$ (= total number of flows $\leq N(N-1)$), $z_{APR} \leftarrow 1, ape \leftarrow 0$
 5. While ($mPaths \geq |\mathcal{F}_s| - 1$) do
 - a. $nCycles \leftarrow 2$
 - b. $nPaths \leftarrow mPaths$
 - c. $\bar{R}_a(s) \leftarrow \bar{R}_e(s)$
 - d. Compute \bar{B} and $\bar{c}^Q, B_{ms}, s \in \mathcal{S}_Q$ or $\bar{c}^B, W_B, s \in \mathcal{S}_B$ for \bar{R}_a

- e. $\min\{B_{ms}\} \leftarrow B_{ms}, s \in \mathcal{S}_Q$ or $\max\{W_B\} \leftarrow W_B, s \in \mathcal{S}_B$
 - f. *While* ($nPaths > 0$) *do*
 - i. Compute and order the values of the function $\xi(f_s)$ – see (7)
 - ii. Find the $nPaths$ with lower value of $\xi(f_s)$
 - iii. Compute with MMRA-S2 new candidate paths for the corresponding O-D pairs and define a new set of first and second choice paths for the service s , $\bar{R}_a(s)$, according to the rules established in [8, Appendix C].
 - iv. Compute \bar{B} and $B_{ms}, B_{Ms}, s \in \mathcal{S}_Q$ or $W_B, s \in \mathcal{S}_B$ for \bar{R}_a
 - v. If $s \in \mathcal{S}_Q$ then
 - A. If ($B_{ms} < \min\{B_{ms}\}_{ini}$ and $B_{Ms} < \min\{B_{Ms}\}_{ini}$) then
 - Compute $W_Q, B_{Mm|Q}$
 - If $W_Q > \max\{W_Q\}$ and $B_{Mm|Q} < \min\{B_{Mm|Q}\}$ then
 - $\min\{B_{ms}\}_{ini} \leftarrow B_{ms}, \min\{B_{Ms}\}_{ini} \leftarrow B_{Ms}$
 - $\max\{W_Q\} \leftarrow W_Q, \min\{B_{Mm|Q}\} \leftarrow B_{Mm|Q}$
 - $\bar{R}_*(s) \leftarrow \bar{R}_a(s)$
 - B. If ($B_{ms} < \min\{B_{ms}\}$) then
 - $\min\{B_{ms}\} \leftarrow B_{ms}$
 - C. Otherwise go to 5.f.vii
 - vi. Otherwise ($s \in \mathcal{S}_B$)
 - A. If ($W_B > \max\{W_B\}_{ini}$) then
 - Compute $W_Q, B_{Mm|Q}$
 - If $W_Q > \max\{W_Q\}$ and $B_{Mm|Q} < \min\{B_{Mm|Q}\}$ then
 - $\max\{W_B\}_{ini} \leftarrow W_B$
 - $\max\{W_Q\} \leftarrow W_Q, \min\{B_{Mm|Q}\} \leftarrow B_{Mm|Q}$
 - $\bar{R}_*(s) \leftarrow \bar{R}_a(s)$
 - B. If ($W_B > \max\{W_B\}$) then
 - $\max\{W_B\} \leftarrow W_B$
 - C. Otherwise go to 5.f.vii
 - vii. (Update $nPaths$)
 - A. $nPaths \leftarrow nPaths - 1$
 - B. If ($nPaths = 0$ and $nCycles = 2$) then
 - $nCycles \leftarrow nCycles - 1$
 - $nPaths \leftarrow |\mathcal{F}_s|$
 - C. Compute \bar{B} and $\bar{c}^Q, s \in \mathcal{S}_Q$ or $\bar{c}^B, s \in \mathcal{S}_B$ for \bar{R}_a
 - D. If ($nPaths \leq 10$ and $ape \geq 1$) then
 - $z_{APR} \leftarrow nPaths \cdot 0.1$
 - E. Otherwise $z_{APR} \leftarrow 1$
- End of the cycle *While* ($nPaths$)
- g. $ape \leftarrow ape + 1$
 - h. If ($ape > 1$) then
 - i. $mPaths \leftarrow mPaths - 1$
- End of the cycle *While* ($mPaths$)
6. $\bar{R}_a(s) \leftarrow \bar{R}_*(s)$

End of the cycle *For* (*s*)

IX. If $W_Q^o > \max\{W_Q\}$ or $B_{Mm|Q}^o < \min\{B_{Mm|Q}\}$ then

1. The best solution is \bar{R}_o

X. Otherwise, the best solution is \bar{R}_*

XI. Compute the objective function values for the best solution

References

1. C. H. Antunes, J. Craveirinha, J. Clímaco, C. Barrico: A multiple objective routing approach for integrated communication networks. In D. Smith, P. Key (eds.), *Proceedings of the 16th International Teletraffic Congress (ITC16) – Teletraffic Engineering in a Competitive World*, pp. 1291–1300. Elsevier (1999)
2. D. Awduche, A. Chiu, A. Elwalid, I. Widjaja, X. Xiao: Overview and principles of Internet traffic engineering. RFC 3272, Network Working Group (2002)
3. D. Awduche, J. Malcolm, J. Agogbua, M. O’Dell, J. McManus: Requirements for traffic engineering over MPLS. RFC 2702, Network Working Group (1999)
4. J. Clímaco, J. Craveirinha: Multicriteria analysis in telecommunication network planning and design – Problems and issues. In J. Figueira, S. Greco, M. Ehrgott (eds.), *Multiple Criteria Decision Analysis – State of the Art Surveys*, volume 78 of *International Series in Operations Research & Management Science*, pp. 899–951. Springer (2005)
5. J. C. N. Clímaco, J. M. F. Craveirinha, M. M. B. Pascoal: Multicriteria routing models in telecommunication networks – Overview and a case study. In Y. Shi, D. L. Olson, A. Stam (eds.), *Advances in Multiple Criteria Decision Making and Human Systems Management: Knowledge and Wisdom*, pp. 17–46. IOS Press (2007)
6. J. C. N. Clímaco, E. Q. V. Martins: A bicriterion shortest path algorithm. *European Journal of Operational Research* **11**(4) (1982) pp. 399–404
7. J. Craveirinha, R. Girão-Silva, J. Clímaco: A meta-model for multiobjective routing in MPLS networks. *Central European Journal of Operations Research* **16**(1) (2008) pp. 79–105
8. J. Craveirinha, R. Girão-Silva, J. Clímaco, L. Martins: *A hierarchical multiobjective routing model for MPLS networks with two service classes – Analysis and resolution approach*. Research Report 5/2007 (ISSN 1645-2631), INESC-Coimbra (www.inescc.pt), Oct. (2007)
9. J. Craveirinha, L. Martins, J. N. Clímaco: Dealing with complexity in a multiobjective dynamic routing model for multiservice networks – A heuristic approach. In *Proceedings of the 15th Mini-EURO Conference on Managing Uncertainty in Decision Support Models (MUDSM 2004)*, Coimbra, Portugal, Sep. 22-24 (2004)
10. J. Craveirinha, L. Martins, T. Gomes, C. H. Antunes, J. N. Clímaco: A new multiple objective dynamic routing method using implied costs. *Journal of Telecommunications and Information Technology* **3** (2003) pp. 50–59
11. H. M. Elsayed, M. S. Mahmoud, A. Y. Bilal, J. Bernussou: Adaptive alternate-routing in telephone networks: Optimal and equilibrium solutions. *Information and Decision Technologies* **14** (1988) pp. 65–74
12. S. C. Erbas: Utilizing evolutionary algorithms for multiobjective problems in traffic engineering. In W. Ben-Ameur, A. Petrowski (eds.), *Proceedings of the International Networks Optimization Conference (INOC 2003)*, pp. 207–212, Evry/Paris, France, Institut National des Télécommunications, Oct. 27-29 (2003)
13. S. C. Erbas, C. Erbas: A multiobjective off-line routing model for MPLS networks. In J. Charzinski, R. Lehnert, P. Tran-Gia (eds.), *Proceedings of the 18th International Teletraffic Congress (ITC-18)*, pp. 471–480, Berlin, Germany, Elsevier, Amsterdam (2003)

14. S. C. Erbas, R. Mathar: An off-line traffic engineering model for MPLS networks. In B. Bakshi, B. Stiller (eds.), *Proceedings of the 27th Annual IEEE Conference on Local Computer Networks (27th LCN)*, pp. 166–174, Tampa, Florida, IEEE Computer Society, Nov. (2002)
15. A. Faragó, S. Blaabjerg, L. Ast, G. Gordos, T. Henk: A new degree of freedom in ATM network dimensioning: Optimizing the logical configuration. *IEEE Journal on Selected Areas in Communications* **13**(7) (1995) pp. 1199–1206
16. J. Granat, A. P. Wierzbicki: Multicriteria analysis in telecommunications. In *Proceedings of the 37th Hawaii International Conference on System Sciences*, Jan. 5-8 (2004)
17. F. P. Kelly: Notes on effective bandwidths. In F. P. Kelly, S. Zachary, I. Ziedins (eds.), *Stochastic Networks: Theory and Applications*, volume 4 of *Royal Statistical Society Lecture Notes Series*, pp. 141–168. Oxford University Press (1996)
18. F. P. Kelly: Routing in circuit-switched networks: Optimization, shadow prices and decentralization. *Advances in Applied Probability* **20**(1) (1988) pp. 112–144
19. F. P. Kelly: Routing and capacity allocation in networks with trunk reservation. *Mathematics of Operations Research* **15**(4) (1990) pp. 771–793
20. J. Knowles, M. Oates, D. Corne: Advanced multi-objective evolutionary algorithms applied to two problems in telecommunications. *BT Technology Journal* **18**(4) (2000) pp. 51–65
21. H. Kochkar, T. Ikenaga, Y. Hori, Y. Oie: Multi-class QoS routing with multiple routing tables. In *Proceedings of IEEE PACRIM'03*, pp. 388–391, Victoria, B.C., Canada, Aug. (2003)
22. F. Kuipers, P. Van Mieghem, T. Korkmaz, M. Krunz: An overview of constraint-based path selection algorithms for QoS routing. *IEEE Communications Magazine* Dec. (2002) pp. 50–55
23. Q. Ma, P. Steenkiste: Supporting dynamic inter-class resource sharing: A multi-class QoS routing algorithm. In *Proceedings of IEEE Infocom'99*, pp. 649–660, New York, Mar. (1999)
24. E. Q. V. Martins, M. M. B. Pascoal, J. L. E. Santos: Deviation algorithms for ranking shortest paths. *International Journal of Foundations of Computer Science* **10**(3) (1999) pp. 247–263
25. L. Martins, J. Craveirinha, J. Clímaco: *A new multiobjective dynamic routing method for multiservice networks – Modeling, resolution and performance*. Research Report 2/2005, INESC-Coimbra (www.inescc.pt), Feb. (2005)
26. L. Martins, J. Craveirinha, J. Clímaco: A new multiobjective dynamic routing method for multiservice networks: Modeling and performance. *Computational Management Science* **3**(3) July (2006) pp. 225–244
27. L. Martins, J. Craveirinha, J. N. Clímaco, T. Gomes: Implementation and performance of a new multiple objective dynamic routing method for multiexchange networks. *Journal of Telecommunications and Information Technology* **3** (2003) pp. 60–66
28. D. Mitra, J. A. Morrison, K. G. Ramakrishnan: ATM network design and optimization: A multirate loss network framework. *IEEE/ACM Transactions on Networking* **4**(4) Aug. (1996) pp. 531–543
29. D. Mitra, J. A. Morrison, K. G. Ramakrishnan: Optimization and design of network routing using refined asymptotic approximations. *Performance Evaluation* **36-37** (1999) pp. 267–288
30. D. Mitra, K. G. Ramakrishnan: Techniques for traffic engineering of multiservice, multipriority networks. *Bell Labs Technical Journal* **6**(1) Jan. (2001) pp. 139–151
31. M. Pióro, D. Medhi: *Routing, Flow, Capacity Design in Communication and Computer Networks*. Morgan Kaufmann Publishers, 1st edition (2004)
32. J. M. Rodrigues, J. C. Clímaco, J. R. Current: An interactive bi-objective shortest path approach: Searching for unsupported nondominated solutions. *Computers & Operations Research* **26** (1999) pp. 789–798
33. E. Rosen, A. Viswanathan, R. Callon: Multiprotocol label switching architecture. RFC 3031, Network Working Group, Jan. (2001)