Hierarchical Multiobjective Routing Model in MPLS Networks with Two Service Classes – A Comparison Case Study

Rita Girão-Silva[a,b], José Craveirinha[a,b], and João Clímaco[b]

[a] Department of Electrical Engineering and Computers, University of Coimbra, Coimbra, Portugal
[b] Institute of Computers and Systems Engineering of Coimbra, University of Coimbra, Coimbra, Portugal

Abstract—A two-level hierarchical multicriteria routing model for multiprotocol label switching networks with two service classes (QoS, i.e., with quality of service requirements, and best effort services) and alternative routing is reviewed in this paper. A heuristic resolution approach, where non-dominated solutions are obtained throughout the heuristic run and kept in an archive for further analysis is also reviewed. In this paper, an extensive analysis of the application of this procedure to two reference test networks for various traffic matrices is presented. Also a comparison of the results of our method with a lexicographic optimization approach based on a multicommodity flow formulation using virtual networks is carried out. Finally, results of a stochastic discrete event simulation model developed for these networks will be shown to illustrate the effectiveness of the resolution approach and to assess the inaccuracies of the analytic results.

Keywords—multiobjective optimization, routing models, simulation, telecommunication networks.

1. Introduction and Motivation

The routing calculation and optimization problems in modern multiservice networks are quite challenging, as the performance requirements in these networks are multidimensional, complex and sometimes contradictory. Routing problems in communication networks consist of the selection of a sequence of network resources (i.e., paths or routes) that will seek the optimization of some objective functions (o.f.), while satisfying a set of constraints. According to the route related metrics that are chosen, the performance of different routing decisions may be measured and quantified.

There are different classes of traffic with different service requirements in multiservice networks. With multiple and heterogeneous QoS (quality of service) routing requirements being taken into account, the routing models are designed to calculate and select one (or more) sequence of network resources (routes), with the aim of seeking the optimization of route related objectives and satisfying certain QoS constraints. There are potential advantages in formulating routing problems in these types of networks as mul-

tiple objective optimization problems, because the trade-offs among distinct performance metrics and other network cost function(s) (potentially conflicting) can be achieved in a consistent manner.

An in-depth methodological discussion of applications of multicriteria analysis in telecommunications seen from a knowledge theory broad perspective, is in [1], while [2] proposes a systematized conceptual framework for multiple criteria routing in QoS/IP networks, using a reference point-based approach.

The authors have presented a meta-model for hierarchical multiobjective network-wide routing optimization in MPLS networks in [3], along with a discussion on some key methodological and modeling issues associated with route calculation, and selection in MPLS networks. The application of this routing model framework is adequate to core or metro-core networks with a limited number of nodes. Two different classes of traffic flows are considered in this optimization approach, QoS (regarded as first priority flows) and BE – best effort (regarded as second priority flows). While the QoS flows have a guaranteed QoS level, related to the required bandwidth, the BE flows are routed with the best possible quality of service but without deteriorating the QoS of the QoS traffic flows. With this approach, the different traffic flows are treated according to their specific features. The routing model considered here is hierarchical, with two different priority levels. In the first priority level, the o.f. are concerned with network level objectives of QoS flows; in the second priority level, the o.f. are related to performance metrics for the different types of QoS services and to a network level objective for the BE traffic flows.

A heuristic approach (HMOR-S2$_{PAS}$ – hierarchical multiobjective routing considering 2 classes of service with a Pareto archive strategy) devised to find "good" solutions (in the sense of multiobjective optimization[1]) to this hierarchical multiobjective routing optimization problem was

---

[1]In multiobjective optimization problems, see [4], one seeks to find non-dominated solutions since optimal (ideal) solutions are usually unfeasible. A non-dominated solution can be defined as a feasible solution such that (in minimization problems) it is not possible to decrease the value of an o.f. without increasing the value of at least one of the other o.f.

proposed in [5]. Its application to two reference test networks, $\mathcal{M}$ and $\mathcal{E}$, used in a benchmarking case study for various traffic matrices was also described. The evaluation of the performance of the proposed heuristic, by using an analytical model and stochastic discrete-event simulation was presented.

In this work, the heuristic approach HMOR-S2$_{PAS}$ is applied to two new networks, denoted by $\mathcal{G}$ and $\mathcal{H}$, obtained by a transformation of an original network in [6], by a redimensioning of the links. An extensive analysis of the application of this procedure to these networks for various traffic matrices is presented. A major objective of this experimental study is to test the developed routing method in new networks with different structure and increased connectivity, as compared with the ones in [5]. Furthermore, the results were obtained in these networks, using analytic and stochastic discrete-event simulation models in order to confirm the effectiveness of this heuristic approach to route calculation and selection in multiservice networks, and to assess the inaccuracies of the analytic results.

Furthermore, for comparison purposes we also implemented a network-wide optimization routing method based on a MCF (multicommodity flow) programming approach with two-path traffic splitting, using lexicographic optimization for dealing with the two main o.f. associated with QoS and BE traffic. This routing method (designated hereafter as MCF-lex-W) is a particular variant of the one proposed in [7] and from our point of view, this type of approach (among the ones previously developed) can be broadly comparable in terms of underlying objectives to our approach. This type of alternative method uses the concept of virtual residual networks whereby, in a first step, the routing calculation is performed for the QoS traffic (seeking to optimize a relevant o.f.) and in a second step the routing calculation for the BE traffic is performed considering only the remaining capacity in the links (resulting from the occupation of the QoS flows). This results in a virtual residual network and it is a classical form of dealing with routing problems in networks with two classes of services of different priority, as in [8], [9]. Since this type of models using MCFs assume deterministic flows (this is an intrinsic limitation of this type of approaches), the comparison with the results of our multiobjective model requires an adaptation to a stochastic environment of the type proposed in [7] and adapted to the developed models as described in Subsection 3.3.

The paper is organized as follows: the two-level hierarchical multiobjective alternative routing model with two service classes is reviewed in Section 2. The main features of the heuristic resolution approach are also reviewed. In the following section, after an explanation on the application of the model to a network case study and the description of the test networks considered in the experimental study, the MCF-lex-W method used for comparison purposes is described. Still in Section 3 the results obtained with this procedure by using analytic results and discrete-event stochastic simulations, for the two new test networks,

considering three load scenarios are presented. The paper ends with a section on conclusions and an outline of future work.

2. Review of the Multiobjective Routing Model and the Heuristic Resolution Approach

In this section we will make a review of the essential aspects of the multiobjective routing model and of the heuristic resolution approach. Due to the complex nature of the model and of the resolution approach, we refer the readers to further details in [3].

2.1. The Multiobjective Routing Model

The model described here is an application of the multiobjective modeling framework (or "meta-model") for MPLS networks proposed in [3]. In this model, two classes of services are considered: QoS and BE. The sets $\mathcal{S}_Q$ and $\mathcal{S}_B$ include the different service types of each class, that may differ in important attributes, namely the required bandwidth.

The network is represented in this model through a capacitated directed graph, with an assigned capacity of C_k to every arc (or 'link') $k \in \mathcal{A}$. The traffic flows are represented in a stochastic form, based on the use of the concept of effective bandwidth[2] for macro-flows and on a generalized Erlang model for estimating the blocking probabilities in the arcs, as in the model used in [12].

A traffic flow is specified by $f_s = (i, j, \overline{\gamma}_s, \overline{\eta}_s)$ for $s \in \mathcal{S} = \mathcal{S}_Q \cup \mathcal{S}_B$ and a stochastic process (usually, a marked point process) is assigned to it. This process describes the arrivals and basic requirements of micro-flows[3], originated at the MPLS ingress node i and destined for the MPLS egress node j, using some LSP (label switched path). The characteristics of the traffic flows are expressed by $\overline{\gamma}_s$, the vectors of traffic engineering attributes of flows of service type s, and by $\overline{\eta}_s$, the vectors containing the description of mechanisms of admission control to all arcs k in the network by calls of flow f_s. The traffic engineering attributes associated with f_s calls and all the links, which may be used by f_s, including priority features, include information on the required effective bandwidth d_s and the mean duration $h(f_s)$ of each micro-flow in f_s.

---

[2]The effective bandwidth can be defined (see [10]) as the minimum amount of bandwidth that can be assigned to a flow or traffic aggregate in order to deliver 'acceptable service quality' to the flow or traffic aggregate. This concept may be used to approximate nodal behavior at the packet level and simplify the analysis at the connection level. Kelly [11] developed a formal mathematical definition of effective bandwidth in a network with stochastic traffic sources and statistical multiplexing. According to this definition, the effective bandwidth can be viewed as a specific stochastic measure of the utilization of transmission network resources by certain packet flow(s). With this concept, the traffic behavior at packet level may be "encapsulated" in a simplified manner.

[3]A micro-flow corresponds in this model to a 'call', that is, a node to node connection request with certain traffic engineering features.

The hierarchical multiobjective routing optimization model considered here has two levels with several o.f. in each level. At the first level, the first priority o.f. include W_Q, the total expected network revenue associated with QoS traffic flows, and $B_{Mm|Q}$, the worst average performance among QoS services, represented by the maximal average blocking probability among all QoS service types. These o.f. are formulated at the network level for the QoS traffic. At the second level, the second priority o.f. include $B_{ms|Q}$, the mean blocking probabilities for flows of type $s \in \mathscr{S}_Q$, and $B_{Ms|Q}$, the maximal blocking probability defined over all flows of type $s \in \mathscr{S}_Q$, as well as the total expected network revenue associated with BE traffic flows, W_B. The o.f. related to blocking probabilities in this second level are average performance metrics of the QoS traffic flows associated with the different types of QoS services. At both levels of optimization, 'fairness' objectives are explicitly considered in the form of min-max objectives: $\min_{\overline{R}}\{B_{Mm|Q}\}$ at the first level, and $\min_{\overline{R}}\{B_{Ms|Q}\}, \forall s \in \mathscr{S}_Q$ at the second level.

Hence the considered two-level hierarchical optimization problem for two service classes P-M2-S2 ('problem – multiobjective with 2 optimization hierarchical levels – with 2 service classes') is:

Problem P-M2-S2

- 1st level $\begin{cases} \text{QoS: Network obj.} & \max_{\overline{R}}\{W_Q\}, \\ & \min_{\overline{R}}\{B_{Mm|Q}\}; \end{cases}$

- 2nd level $\begin{cases} \text{QoS: Service obj.} & \min_{\overline{R}}\{B_{ms|Q}\}, \\ & \min_{\overline{R}}\{B_{Ms|Q}\}, \\ & \forall s \in \mathscr{S}_Q, \\ \text{BE: Network obj.} & \max_{\overline{R}}\{W_B\}; \end{cases}$

subject to equations of the underlying traffic model, with

$$W_{Q(B)} = \sum_{s \in \mathscr{S}_{Q(B)}} A_s^c w_s, \tag{1}$$

$$B_{Mm|Q} = \max_{s \in \mathscr{S}_Q}\{B_{ms}\}, \tag{2}$$

$$B_{ms|Q} = \frac{1}{A_s^o} \sum_{f_s \in \mathscr{F}_s} A(f_s)B(f_s), \tag{3}$$

$$B_{Ms|Q} = \max_{f_s \in \mathscr{F}_s}\{B(f_s)\}, \tag{4}$$

where A_s^o is the total traffic offered by flows of type s, A_s^c is the carried traffic for service type s, $A(f_s)$ is the mean traffic offered associated with f_s (in Erlang), $B(f_s)$ is the node to node blocking probability for all flows f_s, and w_s is the expected revenue per call of service type s. For further details on the calculation of these o.f. see [3].

There are possible conflicts between the o.f. in P-M2-S2. In fact, in many routing situations, the maximization of W_Q may cause a deterioration on some $B(f_s), s \in \mathscr{S}_Q$, for certain traffic flows $A(f_s)$ with low intensity, which tends to increase $B_{Ms|Q}$ and $B_{ms|Q}$, and consequently $B_{Mm|Q}$. This justifies the interest and potential advantage in using multiobjective formulations in this context.

It is important to remark that in the formulation of P-M2-S2, W_Q is a first priority o.f. (together with $B_{Mm|Q}$), while W_B is a second level o.f. This formulation assures that the routing of BE traffic, in a quasi-stationary situation, will not be made at the expense of a decrease in QoS traffic revenue or of an increase in the maximal blocking probability of QoS traffic flows.

The traffic modeling approach used in the routing model is fully described in [3]. In the framework of the basic teletraffic model considered here, the blocking probabilities B_{ks}, for micro-flows of service type s in link k, are calculated by

$$B_{ks} = \mathscr{B}_s\left(\overline{d_k},\overline{\rho_k},C_k\right), \tag{5}$$

with $\mathscr{B}_s$ representing the basic function (implicit in the teletraffic analytical model) that expresses the marginal blocking probabilities, B_{ks}, in terms of $\overline{d_k} = (d_{k1},\ldots,d_{k|\mathscr{S}|})$ (vector of equivalent effective bandwidths d_{ks} for all service types), $\overline{\rho_k} = (\rho_{k1},\ldots,\rho_{k|\mathscr{S}|})$ (vector of reduced traffic loads ρ_{ks} offered by flows of type s to k) and the link capacity C_k. For simplifying purposes, the links are modeled through a multidimensional Erlang system with multirate Poisson traffic inputs. With this type of approximation, the calculation of $\{B_{ks}\}$ can be performed through efficient and robust numerical algorithms, which are essential in a network-wide routing optimization model of this type, for tractability reasons. The classical Kaufman (or Roberts) algorithm [13], [14] was used to calculate the functions $\mathscr{B}_s$ for small values of C_k; for larger values of C_k, approximations based on the uniform asymptotic approximation (UAA) [15] were used, having in mind its efficiency.

The decision variables $\overline{R} = \cup_{s=1}^{|\mathscr{S}|} R(s)$ represent the network routing plans, that is, the set of all the feasible routes (i.e., node to node loopless paths) for all traffic flows, with $R(s) = \cup_{f_s \in \mathscr{F}_s} R(f_s), s \in \mathscr{S}_Q \cup \mathscr{S}_B$ and $R(f_s) = (r^p(f_s)), p = 1,\ldots,M$ with $M = 2$ in this model. An alternative routing principle is used: for each flow f_s the first choice route $r^1(f_s)$ is attempted and if it is blocked the call will try the second choice route $r^2(f_s)$. A request will be blocked only if $r^2(f_s)$ is also blocked.

This routing optimization approach is of network-wide type, which means that the main o.f. of a given service class depend explicitly on all traffic flows in the network. Therefore, a full representation of the relations between the o.f. is achieved, taking into account the interactions between the multiple traffic flows associated with different services. This is accomplished by the features of the traffic model used to obtain the blocking probabilities $B(f_s)$, as the contributions of all traffic flows, which may use every link of the network are considered according to the approach in [3]. The focus is on the routing optimization from a global perspective (i.e., considering an explicit representation of all the traffic flows in the network and their interactions), which is the closest to reality. This is a major difference in comparison with other routing models that have been proposed for networks with two service classes, based on some form of decomposition of the network

representation, leading to the consideration of 'virtual networks', one for each service class (e.g. in [7]).

The routing problem P-M2-S2 is highly complex, mainly because of two factors: all o.f. are strongly interdependent (via the $\{B(f_s)\}$), and all the o.f. parameters and (discrete) decision variables $\overline{R}$ (network route plans) are also interdependent. All these interdependencies are defined explicitly or implicitly through the underlying traffic model. Even in a simplest degenerate case, considering single service with single-criterion optimization and no alternative routing, the problem is NP-complete in the strong sense (see [16]). Considering the form of P-M2-S2, one may conclude on the great intractability of this problem.

2.2. The Heuristic Resolution Approach

The heuristic procedure HMOR-S2$_{PAS}$ (fully described in [5] and references therein) used to solve (in a multi-criteria analysis sense) the routing problem P-M2-S2 is reviewed here. Using the theoretical foundations described in [17], this heuristic is based on the recurrent calculation of solutions to an auxiliary constrained bi-objective shortest path problem $\mathscr{P}_{s2}^{(2)}$, formulated for every end-to-end flow f_s,

$$\min_{r(f_s) \in \mathscr{D}(f_s)} \left\{ m^n(r(f_s)) = \sum_{k \in r(f_s)} m_{ks}^n \right\}_{n=1;2} .$$

The path metrics m^n to be minimized are the marginal implied costs[4] $m_{ks}^1 = c_{ks}^{Q(B)}$ and the marginal blocking probabilities $m_{ks}^2 = -\log(1 - B_{ks})$; $\mathscr{D}(f_s)$ is the set of all feasible loopless paths for flow f_s, satisfying specific traffic engineering constraints for flows of type s. The efficiency of different candidate routes can be compared, considering both path metrics: the loss probabilities experienced along the candidate routes and the knock-on effects upon the other routes in the network (effects related to the acceptance of a call on that given route). It is important to remark that these network metrics are associated with the first level o.f. of P-M2-S2: the minimization of the metric blocking probability tends, at a network level, to minimize the maximal node-to-node blocking probabilities $B(f_s)$, while the minimization of the metric implied cost tends to maximize the total average revenue W_T.

In the heuristic, the auxiliary constrained shortest path problem $\mathscr{P}_{s2}^{(2)}$ is solved by the algorithm MMRA-S2 (modified multiobjective routing algorithm for multiservice networks, considering 2 classes of service) described in [18]. Generally, there is no feasible solution minimizing the two o.f. simultaneously. Therefore, the aim of the resolution of this problem is finding a 'best' compromise path from the set of non-dominated solutions, according to

a system of preferences embedded in the working of the algorithm MMRA-S2. The implementation of this system of preferences relies on the definition of preference regions in the o.f. space obtained from aspiration and reservation levels (preference thresholds), defined for the two o.f.

The generation and selection of candidate solutions $(r^1(f_s), r^2(f_s))$ by MMRA-S2 for each f_s is based on rules that consider the network topology and the need to make a distinction between real time and non-real time QoS services, and BE services. An instability phenomenon may arise in the path selection procedure, as shown by a theoretical analysis of the model and confirmed by extensive experimentation: the route sets $\overline{R}$ (obtained by successive application of MMRA-S2 to every flow f_s) often tend to oscillate between certain solutions, some of which may lead to poor global network performance under the prescribed metrics. To avoid this instability, not all the paths of all the flows are liable to change on each iteration. A set of candidate paths for possible routing improvement are chosen by increasing order of a function $\xi(f_s)$ of the current $(r^1(f_s), r^2(f_s))$, as proposed in [18]. With this function $\xi(f_s)$ preference (concerning the calculation of new routes) is given to the flows, for which the route $r^1(f_s)$ has a low implied cost, and the route $r^2(f_s)$ has a high implied cost or to the flows, which currently have worse end-to-end blocking probability. A variation on the selected paths is performed, leaving the others unaltered.

In the dedicated heuristic HMOR-S2, each new solution is obtained by 'processing' the current best solution: routing solutions $R(s)$ for each service $s \in \mathscr{S}$ are sought which dominate the current one in terms of the so-called o.f. of interest for the service (the first level o.f. and the second level o.f. $B_{ms|Q}$ and $B_{Ms|Q}$ if $s \in \mathscr{S}_Q$, or W_B if $s \in \mathscr{S}_B$). This strategy leads to strict limitations being imposed on the acceptance of a new solution, and consequently some interesting solutions to the routing problem may not be further pursued. Therefore, instead of simply discarding every solution that does not dominate the current one, we have devised the PAS variant where some possibly interesting solutions are stored throughout the execution of the heuristic, and later checked in order to try and find the "best" possible solution to the problem in hand. The management rules of the archive (that is, addition and removal of solutions from the archive) and the evaluation of the solutions stored in the archive after the end of the outer cycle of the algorithm (in order to choose the "best" possible solution to the problem under analysis) are fully described in [5].

The analysis of the solutions stored in the archive relies on a system of priority regions in the bidimensional o.f. space, defined by preference thresholds (*requested* (or aspirational) and *acceptable* (or reservation) thresholds for each network function W_Q and $B_{Mm|Q}$). As an example of the definition of priority regions in the bidimensional o.f. space of the solutions in the archive, see Fig. 1.

The ideal optimum is represented by $O*$ and is obtained when both first level o.f. W_Q and $B_{Mm|Q}$ are optimized.

---

[4]The *marginal implied cost* for QoS(BE) traffic, $c_{ku}^{Q(B)}$, associated with the acceptance of a connection (or "call") of traffic f_u of any service type $u \in \mathscr{S}$ on a link k represents the expected value of the traffic loss induced on all QoS(BE) traffic flows resulting from the capacity decrease in link k (see [17]).
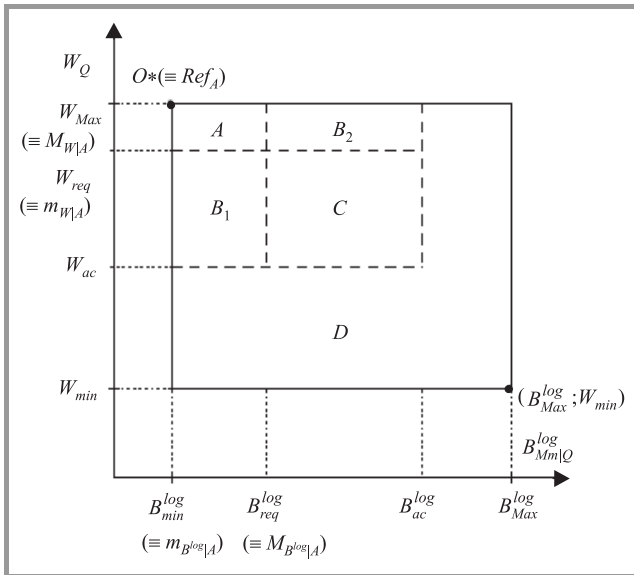
Fig. 1. QoS requirements used to define priority regions in the bidimensional o.f. space.

The region, for which the requested levels are satisfied for both o.f. is the first priority region A; the regions, for which only one of the requested values is satisfied and an acceptable value is guaranteed for the other metric are the second priority regions B_1 and B_2 (note that B_2 will be considered preferable to B_1 because, for solutions in any second priority region, preference is given to the one with greater W_Q even if with greater $B_{Mm|Q}$); the region where only acceptable values are guaranteed for both metrics is the third priority region C. Beyond the acceptable values, there lies the least priority region D. The preference thresholds used to define the priority regions are calculated in a fully automated manner (see [5]).

The approach chosen to select the "best" solution in the best possible priority region relies on the minimization of a weighted Chebyshev distance to a reference point. In this approach, reference (aspiration and reservation) levels are specified for each criterion. Let $W_{av} = \frac{W_{min}+W_{max}}{2}$, where $W_{min}(W_{max})$ is the minimal(maximal) value of W_Q in all the solutions in the archive, and $B_{av} = \frac{B_{min}+B_{max}}{2}$, where $B_{min}(B_{max})$ is the minimal(maximal) value of $B_{Mm|Q}$ in all the solutions in the archive. The reference levels are defined by $W_{req} = \frac{W_{av}+W_{max}}{2}$ and $W_{ac} = \frac{W_{min}+W_{av}}{2}$ for the QoS traffic revenue and $B_{req}^{log} = -\log\left(1 - \frac{B_{min}+B_{av}}{2}\right)$ and $B_{ac}^{log} = -\log\left(1 - \frac{B_{av}+B_{max}}{2}\right)$ for the blocking probability $B_{Mm|Q}$. The weighted Chebyshev distance of a non-dominated solution in a given preference region to the associated aspiration point is calculated, and the "best" solution will be the one in the best possible priority region that minimizes that distance.

Defining $\mathscr{R}$ as the best possible priority region in the o.f. space where at least one solution ρ can be found, a specific reference point $\left(\mathscr{C}_{1|\mathscr{R}}^{*};\mathscr{C}_{2|\mathscr{R}}^{*}\right)$ can be chosen in $\mathscr{R}$ as the ideal point in that region. The ideal point in each rect-

angular region is the top left corner of that region. As an example, see the reference point for region A (Ref_A) in Fig. 1. For a non-rectangular region such as D, the reference point is the ideal point of the whole o.f. space $O*$. Other parameters that must be defined are the minimum $m_{i|\mathscr{R}}$ and maximum $M_{i|\mathscr{R}}$ values of each metric i for region $\mathscr{R}$. As an example, see the minimum and maximum values for both metrics in region A in Fig. 1.

The problem of selection of the final solution considers a weighted Chebyshev norm:

$$\min_{\rho\in\mathscr{R}} \max_{i=1,2}\left\{ w_{i|\mathscr{R}}\left|\mathscr{C}_i(\rho)-\mathscr{C}_{i|\mathscr{R}}^{*}\right|\right\},$$

where $\mathscr{C}_1(\rho) = B_{Mm|Q}^{log}(\rho)$ and $\mathscr{C}_2(\rho) = W_Q(\rho)$ are the metrics for solution ρ. The weights in the weighted Chebyshev distance, $w_{i|\mathscr{R}} = \frac{1}{M_{i|\mathscr{R}}-m_{i|\mathscr{R}}}$, allow the Chebyshev metrics $\left\{ w_{i|\mathscr{R}}\left|\mathscr{C}_i(\rho)-\mathscr{C}_{i|\mathscr{R}}^{*}\right|\right\}$ to be dimension free and proportional to the size of the rectangular region. This weighted Chebyshev norm is more adequate to the adopted technique of search and selection of non-dominated solutions in rectangular preference regions. In fact, the use of the weights (as defined in the method) makes the contour of the rectangle a isocost Chebyshev line for each particular region.

3. Experimental Results

3.1. Application of the Model to a Network Case Study

The network case study considered here is obtained from changes on the network models in [7] and [6]. An overview of the relevant features of the model proposed in this reference is provided here for a better understanding of the case study.

In [7], a model is proposed for traffic routing and admission control in multiservice, multipriority networks supporting traffic with different QoS requirements. Deterministic models are used in the calculation of paths, in particular mathematical programming models based on MCFs, rather than stochastic traffic models. The MCF models are only a rough approximation in this context and, in fact, they tend to under-evaluate the blocking probabilities. Therefore, the authors of [7] propose an adaptation of the original model, so as to obtain 'corrected' models, which provide a better approximation in a stochastic traffic environment. A simple technique to adapt the MCF model to a stochastic environment is the compensation of the requested values of the flows bandwidths in the MCF model with a factor $\alpha \geq 0.0$. With this compensation technique, the effect of the random fluctuations of the traffic that are typical of stochastic traffic models can be modeled. 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. In the application example in [7], three values of α are proposed: $\alpha = 0.0$ corresponds to the deterministic approach; $\alpha = 0.5$ is the compensation parameter 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 o.f. of the routing problems in [7] are the revenues W_Q and W_B, associated with QoS and BE flows, which should be maximized. A bi-criteria lexicographic optimization formulation is considered, 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 [7], a base matrix $T = [T_{ij}]$ with offered bandwidth values from node i to node j [Mbit/s] is given. A multiplier $m_s \in [0.0; 1.0]$ with $\sum_{s \in \mathscr{S}} m_s = 1.0$ is applied to these matrix values to obtain the offered bandwidth of each flow f_s of service type s to the network. In our stochastic traffic model, a matrix of offered traffic $A(f_s)$ is obtained by transforming the base matrix T:

$$A(f_s) \approx \frac{m_s T_{ij}}{d_s u_0} - \alpha \sqrt{\frac{m_s T_{ij}}{d_s u_0}} \text{ [Erl]}, \qquad (6)$$

if $\frac{m_s T_{ij}}{d_s u_0} > \alpha^2$ and both $T(f_s) = m_s T_{ij}$ and $A(f_s)$ are high. Otherwise,

$$A(f_s) \approx \frac{m_s T_{ij}}{d_s u_0} \text{ [Erl]}, \qquad (7)$$

where u_0 is a basic unit of transmission [bit/s].

In the original traffic routing model in [7], traffic splitting is used. This technique is not used in the model considered here.

3.2. Application of the Model to Two Different Test Networks

The routing model was applied to the test networks $\mathscr{G}$ and $\mathscr{H}$, for which the topology is depicted in Fig. 2. It has $|\mathscr{N}| = 10$ nodes, with 16 pairs of nodes linked by a direct arc and a total of $|\mathscr{A}| = 32$ unidirectional arcs, which means their average node degree is $\delta = 3.2$. As their average node degree is higher, these two networks $\mathscr{G}$ and $\mathscr{H}$ have more connectivity than networks $\mathscr{M}$ and $\mathscr{E}$ ($\delta_{\mathscr{M}} = 2.5$
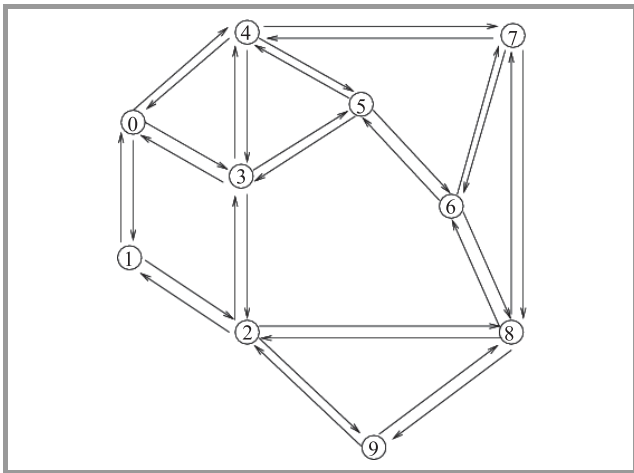

Fig. 2. Network topology for test networks $\mathscr{G}$ and $\mathscr{H}$ [6].

and $\delta_{\mathscr{E}} = 2.4$), studied in [5]. Each bandwidth C'_k [Mbit/s] of each arc k for each of the networks is shown in Tables 1 and 2, and it was obtained by employing a very simple network dimensioning algorithm, explained below.

The test networks $\mathscr{G}$ and $\mathscr{H}$ were obtained after a redimensioning of the original network $\mathscr{O}$ given in [6]. This network $\mathscr{O}$ has a topology similar to the one in Fig. 2, with a capacity of $C'_k = 50$ Mbit/s for each arc, which is an equivalent to a capacity of $C_k = \frac{C'_k}{u_0} = 3125$ channels, as $u_0 = 16$ kbit/s. The offered traffic matrix is also provided in [6]. A routing solution using only shortest path direct routing, typical of Internet conventional routing algorithms is taken into account. In this routing solution, only one path for each flow (i.e., without an alternative path) is considered. The initial solution is the same for all services $s \in \mathscr{S}$ and the unidirectional paths for any given pair of nodes are symmetrical. The path 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, the one with maximal bottleneck bandwidth (i.e., the minimal capacity of its arcs) is chosen; if there is more than one shortest path with equal bottleneck bandwidth, the choice is arbitrary.

The dimensioning of link capacities was made as follows. A value β_s for the mean blocking probabilities for flows of type s, B_{ms}, is defined with a possible variation of Δ_B. The matrix of offered traffic $A(f_s)$ is obtained from the traffic matrix T in [6] with $\alpha = 0.0$ (the value of α for which the load is higher). Considering the routing solution for network $\mathscr{O}$, the mean blocking probabilities B_{ms} are calculated and compared with the prescribed values at the beginning of the algorithm. If $B_{ms} > \beta_s$ for service s, then the links in paths for flows of service s will have their capacity increased; if $B_{ms} < \Delta_B \beta_s$ for service s, then the links in paths for flows of service s will have their capacity decreased. The algorithm proceeds iteratively until it converges (i.e., $\Delta_B \beta_s < B_{ms} < \beta_s, \forall s \in \mathscr{S}$). In some of the performed experiments, the algorithm oscillated between two different solutions, which prevented it from converging. Therefore, a maximum number of runs was also established, so as to avoid this situation.

The test networks $\mathscr{G}$ and $\mathscr{H}$ were dimensioned using this very simple network dimensioning algorithm, for $\beta_s = 0.1$ and $\beta_s = 0.12$ respectively, with $\Delta_B = 0.9$. This means that a situation of very high blocking, associated with traffic overload for all services, was considered (for $\alpha = 0.0$) in the dimensioning operation. The aim was a comparison of the performance of the considered static routing methods in overload conditions ($\alpha = 0.0$) and in low, and very low blocking conditions for the QoS traffic for $\alpha = 0.5$ and $\alpha = 1.0$. The original network $\mathscr{O}$ was not used in this study because it was dimensioned for extremely low blocking probabilities.

The traffic matrix $T = [T_{ij}]$ with offered total bandwidth values from node i to node j [Mbit/s] provided in [6] is used as an input to the routing model considered here. The routing model and other features proposed by [6] were not taken into account.

Table 1

Bandwidth of each arc C'_k, in Mbit/s, for the test network $\mathscr{G}$

↘	0	1	2	3	4	5	6	7	8	9
0		40.64		44.384	40.64					
1	40.64		35.024							
2		35.024		35.024					36.896	38.768
3	44.384		35.024		42.512	38.768				
4	40.64			42.512		44.384		40.64		
5				38.768	44.384		38.768			
6						38.768		46.256	40.64	
7					40.64		46.256		38.768	
8			36.896				40.64	38.768		44.384
9			38.768						44.384	

Table 2

Bandwidth of each arc C'_k, in Mbit/s, for the test network $\mathscr{H}$

↘	0	1	2	3	4	5	6	7	8	9
0		39.6		43.76	39.6					
1	39.6		33.36							
2		33.36		33.36					35.44	37.52
3	43.76		33.36		41.68	37.52				
4	39.6			41.68		43.76		39.6		
5				37.52	43.76		37.52			
6						37.52		45.84	39.6	
7					39.6		45.84		37.52	
8			35.44				39.6	37.52		43.76
9			37.52						43.76	

For both networks, the number of channels C_k is $C_k = \left\lceil \frac{C'_k}{u_0} \right\rceil$, with basic unit capacity $u_0 = 16$ kbit/s. There are $|\mathscr{S}| = 4$ service types with the features displayed in Table 3.

Table 3

Service features on the test networks $\mathscr{G}$ and $\mathscr{H}$

Service	Class	d'_s [kbit/s]	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	9	0.25

The values of the required bandwidth d'_s in kbit/s are also in the table. The expected revenues for calls of type s, w_s, are equal to the required effective bandwidths $d_s = \frac{d'_s}{u_0}$ [channels]: $w_s = d_s, \forall s \in \mathscr{S}$. The average duration of a type s call is h_s and the maximum number of arcs for a type s call is D_s.

3.3. Routing Method Used for Comparison Purposes

Next we describe the MCF-lex-*W* routing model, based on MCFs with lexicographic optimization and considering two-path traffic splitting. This model is based on the one in [7] and it is used as an alternative benchmarking method for comparison with our multiobjective model. From a theoretical point of view, and considering the conceptual framework developed in [3], this type of model is also a network-wide optimization approach with features that make it an adequate alternative method for a 'fair' comparison with our model. The results with the method HMOR-S2$_{PAS}$ considered in this paper are compared with results from this routing procedure MCF-lex-*W*.

In this routing procedure, we first seek to route the QoS traffic flows in the given network. Next, we seek to route the BE traffic flows in a virtual network, where the arcs of the original network have a reduced capacity given by the original arc capacity minus the capacity used in the routing of the QoS flows. In the process of routing calculation, the aim is the maximization of the revenue of QoS and BE carried traffic (represented by the node-to-node offered bandwidth), using a lexicographic optimization approach. Traffic splitting is allowed, in situations where it is advantageous. There is the possibility of dividing the required bandwidth of each flow by multiple paths from source to destination, allowing for a better load distribution in the network. If the network is unable to accommodate all the traffic that is offered, a technique of admission control based on traffic thinning can be used, as proposed in [7].

Considering a traffic flow of the service s represented by f_s, originated at the MPLS ingress node i and destined for the MPLS egress node j, the bandwidth offered by that flow to the network is $T(f_s) = m_s T_{ij}$, as mentioned in Subsection 3.1. For each flow, a set of $L(f_s)$ feasible paths may be obtained, $\mathscr{L}(f_s) = \{p^0(f_s), p^1(f_s), \cdots, p^{L(f_s)-1}(f_s)\}$. Of all the possible paths between i and j, the ones with a number of arcs inferior to D_s (maximal number of arcs established for service s calls) are feasible. In the implemented model, the total bandwidth offered by flow f_s may be divided by $N_L = 2$ of these feasible paths, allowing for the possibility of traffic splitting. Let us define $x^l(f_s)$ as the amount of bandwidth of f_s that will be offered to the l-th path $p^l(f_s)$, and $y^l(f_s)$ as a binary variable, which is equal to 1 if the l-th path is actually used and 0 otherwise. Therefore, the following conditions have to be met, $\forall f_s \in \mathscr{F}_s, s \in \mathscr{S}$:

$$\sum_{l=0}^{L(f_s)-1} x^l(f_s) \leq T(f_s), \tag{8}$$

$$0 \leq x^l(f_s) \leq T(f_s), \quad \forall l = 0, \cdots, L(f_s)-1, \tag{9}$$

$$x^l(f_s) \leq T(f_s) y^l(f_s), \quad \forall l = 0, \cdots, L(f_s)-1, \tag{10}$$

$$\sum_{l=0}^{L(f_s)-1} y^l(f_s) \leq N_L = 2. \tag{11}$$

The o.f. used in this routing method is the maximization of the network revenue $W_T = \sum_{s \in \mathscr{S}} \sum_{f_s \in \mathscr{F}_s} \sum_{l=0}^{L(f_s)-1} w_s x^l(f_s)$ that results from carrying the bandwidth offered by all the traffic flows to all the feasible paths, which are actually used. The possibility of traffic splitting should provide a flexible distribution of the load in the network, so as to maximize the carried traffic. This is particularly relevant in the context of this routing model since, after establishing the optimal routes of the QoS traffic (for which the whole average bandwidth demand is satisfied), it is necessary to calculate the routes for the BE traffic in the virtual residual network, so as to maximize the BE carried traffic.

The type of problem to be solved in this routing procedure is

Problem P-MCF-lex-W$_\mathscr{S}$

$$\max \left\{ \sum_{s \in \mathscr{S}} \sum_{f_s \in \mathscr{F}_s} \sum_{l=0}^{L(f_s)-1} w_s x^l(f_s) \right\}$$

subject to conditions (8)–(11) and

$$v_k \leq C'_k, \forall k \in \mathscr{A},$$

$$v_k = \sum_{s \in \mathscr{S}} \sum_{f_s \in \mathscr{F}_s} \sum_{l=0}^{L(f_s)-1} a^l_k(f_s) x^l(f_s), \forall k \in \mathscr{A},$$

where $a^l_k(f_s)$ is a binary variable equal to 1 if the link k belongs to $p^l(f_s)$, the l-th path for flow f_s, and 0 otherwise. The parameter v_k is the total load carried in each arc $k \in \mathscr{A}$.

The routing calculation approach in the case where QoS and BE traffic classes coexist uses a lexicographic formulation as the one in [7].

Firstly, the problem P-MCF-lex-W$_{\mathscr{S}_Q}$ is solved, and only the QoS traffic is considered. As a result, the values $x^l(f_s), \forall l = 0, \cdots, L(f_s)-1, f_s \in \mathscr{F}_s, s \in \mathscr{S}_Q$ are obtained, which give the amount of bandwidth that is routed in each of the feasible paths for each of the QoS flows. Also, as a result of this problem, an information on v_k is obtained. Let this load be represented by $v_{k(Q)}$.

Secondly, the problem P-MCF-lex-W$_{\mathscr{S}_B}$ is solved, that is, only the BE traffic is considered. In this second problem, a virtual network consisting of the same links but with residual capacities $C'_k - v_{k(Q)}, \forall k \in \mathscr{A}$ is considered. The possibility of BE traffic thinning was considered, as the network has a reduced arc capacity and there is the possibility that not all the BE traffic flows may be carried.

After the resolution of the second problem, the values $x^l(f_s), \forall l = 0, \cdots, L(f_s)-1, f_s \in \mathscr{F}_s, s \in \mathscr{S}_B$ are obtained, which gives us the amount of bandwidth that is routed in each of the feasible paths for each of the BE flows.

The resolution of both problems was performed by CPLEX 12.3.

Once both problems have been solved, the traffic representation model is transformed in order to obtain an approximation suitable to a stochastic traffic environment, hence enabling a comparison with the o.f. values obtained by HMOR-S2. This adaptation is performed as in [7] and considers three different values for the compensation parameter α (see explanation in Subsection 3.1). A matrix of offered traffic in Erlang is obtained by a transformation similar to Eqs. (6)–(7), that is

$$A^l(f_s) \approx \frac{x^l(f_s)}{d_s u_0} - \alpha \sqrt{\frac{x^l(f_s)}{d_s u_0}} \text{ [Erl] if } \frac{x^l(f_s)}{d_s u_0} > \alpha^2,$$

$$A^l(f_s) \approx \frac{x^l(f_s)}{d_s u_0} \text{ [Erl], otherwise}.$$

The arc capacity C'_k in Mbit/s (see Tables 1 and 2) is converted to a capacity of $C_k = \left\lceil \frac{C'_k}{u_0} \right\rceil$ channels. Once the offered traffic in Erlang and the arc capacities in circuits are known, the blocking probability for each offered flow in this stochastic environment may be calculated.

The blocking probabilities B_{ks}, for micro-flows of service type s in link k, are calculated as in Eq. (5). Afterwards, the blocking of each flow along its path is obtained, $B^l(f_s)$. As the offered traffic is also known, the calculation of the o.f. may be performed as in Eqs. (1)–(4). For further details, see Subsection 2.1.

3.4. Analytical Results

An analytical study was performed, where results using just a basic version of the heuristic without storage of current non-dominated solutions, HMOR-S2, were obtained. In these runs of the basic heuristic, the initial solution consists of the shortest path direct routing, typical of Internet

Table 4

Average o.f. values with 95% confidence intervals, for simulations with the routing plan obtained
with the different heuristic strategies in network $\mathscr{G}$

Obj. func.	MCF-lex-W method solution	Routing method proposed by the authors						
		Initial solution	HMOR-S2 (Basis)		HMOR-S2$_{PAS}$(i)		HMOR-S2$_{PAS}$(f)	
			Analytical	Static routing model	Analytical	Static routing model	Analytical	Static routing model
				$\alpha = 0.0$				
W_Q	**20907.71**◁	**20859.85**	**21686.92***	**21686.05**±37.51	**21688.63**◇	**21688.42**±36.97	**21690.16**★	**21690.52**±37.22
$B_{Mm|Q}$	**0.110**	**0.110**	**0.00661**	**0.00756**±0.000848	**0.00595**	**0.00679**±0.00105	**0.00545**	**0.00619**±0.00119
$B_{m1|Q}$	0.110	0.110	0.00661	0.00756±0.000848	0.00595	0.00679±0.00105	0.00545	0.00619±0.00119
$B_{m2|Q}$	0.0636	0.0689	0.000453	0.000892±0.000159	0.000480	0.000881±0.000152	0.000465	0.000828±0.000124
$B_{m3|Q}$	0.00236	0.00308	0.000274	0.000293±2.30·10^{-5}	0.000273	0.000288±1.86·10^{-5}	0.000275	0.000288±2.64·10^{-5}
$B_{M1|Q}$	0.242	0.555	0.0684	0.0677±0.00869	0.0532	0.0653±0.0140	0.0613	0.0771±0.0117
$B_{M2|Q}$	0.147	0.378	0.00302	0.00700±0.00134	0.00756	0.00869±0.00131	0.00699	0.00794±0.00174
$B_{M3|Q}$	0.00622	0.0190	0.00312	0.00316±0.000298	0.00311	0.00317±0.000333	0.00287	0.00288±0.000312
W_B	*6918.87*	*6738.68*	*7167.15*	*7168.36±12.05*	*7163.81*	*7166.23±10.85*	*7158.14*	*7161.10±11.67*
				$\alpha = 0.5$				
W_Q	**17678.97**▷	**17611.81**	**17685.88**†	**17683.53**±15.54	**17685.89**•	**17683.53**±15.54	**17685.89**⊙	**17683.53**±15.54
$B_{Mm|Q}$	**0.00158**	**0.0160**	**1.13·10^{-5}**	**9.90·10^{-7}** ±7.83·10^{-7}	**1.13·10^{-5}**	**9.47·10^{-7}** ±7.67·10^{-7}	**1.04·10^{-5}**	**8.59·10^{-7}** ±8.23·10^{-7}
$B_{m1|Q}$	0.00158	0.0160	1.13·10^{-5}	0	1.13·10^{-5}	0	1.04·10^{-5}	0
$B_{m2|Q}$	0.000864	0.00926	3.3·10^{-9}	0	3.3·10^{-9}	0	7.2·10^{-9}	0
$B_{m3|Q}$	2.68·10^{-5}	0.000371	8.93·10^{-7}	9.90·10^{-7} ±7.83·10^{-7}	8.14·10^{-7}	9.47·10^{-7} ±7.67·10^{-7}	6.25·10^{-7}	8.59·10^{-7} ±8.23·10^{-7}
$B_{M1|Q}$	0.00485	0.147	0.000143	0	0.000143	0	0.000128	0
$B_{M2|Q}$	0.00273	0.0866	1.03·10^{-7}	0	1.03·10^{-7}	0	4.45·10^{-7}	0
$B_{M3|Q}$	9.52·10^{-5}	0.00353	4.52·10^{-6}	2.24·10^{-5} ±1.72·10^{-5}	4.46·10^{-6}	2.24·10^{-5} ±1.72·10^{-5}	4.46·10^{-6}	2.24·10^{-5} ±1.72·10^{-5}
W_B	*5275.03*	*5247.65*	*5296.56*	*5297.19±12.83*	*5296.56*	*5297.19±12.83*	*5296.57*	*5297.18±12.84*
				$\alpha = 1.0$				
W_Q	**16028.11**×	**16025.69**	**16028.14**‡	**16077.61**±15.03	**16028.14**□	**16077.61**±15.03	**16028.14**⊗	**16077.61**±15.03
$B_{Mm|Q}$	**6.45·10^{-6}**	**0.000577**	**5·10^{-10}**	**0**	**5·10^{-10}**	**0**	**5·10^{-10}**	**0**
$B_{m1|Q}$	6.45·10^{-6}	0.000577	5·10^{-10}	0	5·10^{-10}	0	5·10^{-10}	0
$B_{m2|Q}$	3.79·10^{-6}	0.000334	<1·10^{-10}	0	<1·10^{-10}	0	<1·10^{-10}	0
$B_{m3|Q}$	1.00·10^{-7}	1.16·10^{-5}	<1·10^{-10}	0	<1·10^{-10}	0	<1·10^{-10}	0
$B_{M1|Q}$	4.81·10^{-5}	0.00650	1.27·10^{-8}	0	1.27·10^{-8}	0	1.27·10^{-8}	0
$B_{M2|Q}$	2.38·10^{-5}	0.00347	<1·10^{-10}	0	<1·10^{-10}	0	<1·10^{-10}	0
$B_{M3|Q}$	7.62·10^{-7}	0.000123	2·10^{-10}	0	2·10^{-10}	0	2·10^{-10}	0
W_B	*3340.47*	*3354.76*	*3355.88*	*3350.97±24.92*	*3355.88*	*3350.97±24.92*	*3355.88*	*3350.97±24.92*

MCF-lex-W method solution: ◁) 96.29%; ▷) 99.96%; ×) 99.75% of W_Q^{ideal} (the ideal revenue extracted from the data in [6]);
HMOR-S2: *) 99.88%; †) 100%; ‡) 99.75% of W_Q^{ideal}; HMOR-S2$_{PAS}$(i): ◇) 99.89%; •) 100%; □) 99.75% of W_Q^{ideal}; HMOR-S2$_{PAS}$(f):
★) 99.90%; ⊙) 100%; ⊗) 99.75% of W_Q^{ideal}.

conventional routing algorithms, such as the ones used in the network dimensioning algorithm. In further analytical studies, two different types of tests were conducted for the heuristic HMOR-S2$_{PAS}$:

- (i) tests: the initial solution is a solution typical of Internet conventional routing algorithms, such as the one used in the basic version runs.

- (f) tests: the initial solution of the HMOR-S2$_{PAS}$ heuristic is the routing plan obtained at the end of the basic heuristic runs for each specific α. With this experiment, it is possible to check whether that heuristic variant can improve the quality of the final solutions obtained with HMOR-S2 as an alternative to the direct use of the heuristic variant (as in the case of the (i) tests).

The multiobjective routing model in [6] is quite different from the one considered here, so no results concerning any of the o.f. considered here is provided in [6]. The only results that can be extracted from the proposed model in [6] are approximate ideal values for the QoS flows revenue, W_Q^{ideal}. The analytical results concerning the QoS flows revenue W_Q were compared with these approximate ideal values.

The experiments with the HMOR-S2$_{PAS}$ were conducted with an archive of size 5, chosen empirically after extensive experimentation. This experimentation showed that an increase in the archive size would not necessarily lead to better final results, because at the end of the heuristic run, when the final solution is actually chosen from those in the archive, the top 5 solutions tend to be the same regardless of the archive size.

Table 5

Average o.f. values with 95% confidence intervals, for simulations with the routing plan obtained with the different heuristic strategies in network $\mathscr{H}$

Obj. Func.	MCF-lex-W method solution	Routing method proposed by the authors						
		Initial solution	HMOR-S2 (Basis)		HMOR-S2$_{PAS}$(i)		HMOR-S2$_{PAS}$(f)	
			Analytical	Static routing model	Analytical	Static routing model	Analytical	Static routing model
				$\alpha = 0.0$				
W_Q	**20602.28**◁	**20358.90**	**21576.67***	**21559.04**±31.57	**21578.99**◇	**21563.15**±32.06	**21616.01**⋆	**21597.91**±30.13
$B_{Mm\|Q}$	**0.147**	**0.169**	**0.0287**	**0.0306**±0.00147	**0.0299**	**0.0315**±0.00131	**0.0224**	**0.0245**±0.00175
$B_{m1\|Q}$	0.147	0.169	0.0287	0.0306±0.00147	0.0299	0.0315±0.00131	0.0224	0.0245±0.00175
$B_{m2\|Q}$	0.0891	0.111	0.00456	0.00696±0.000526	0.00410	0.00641±0.000370	0.00341	0.00580±0.000523
$B_{m3\|Q}$	0.00346	0.00536	0.00171	0.00170±7.97·10^{-5}	0.00148	0.00147±6.86·10^{-5}	0.000596	0.000601±4.99·10^{-5}
$B_{M1\|Q}$	0.272	0.711	0.150	0.177±0.0194	0.330	0.312±0.0295	0.146	0.157±0.0231
$B_{M2\|Q}$	0.167	0.518	0.0162	0.0289±0.00289	0.0180	0.0304±0.00572	0.0145	0.0215±0.00229
$B_{M3\|Q}$	0.00699	0.0293	0.00828	0.00822±0.000302	0.00964	0.00959±0.000348	0.00362	0.00376±0.000476
W_B	*6724.15*	*6434.17*	*6877.69*	*6886.47±9.03*	*6905.99*	*6914.48±11.01*	*6927.67*	*6935.83±10.55*
				$\alpha = 0.5$				
W_Q	**17670.25**▷	**17419.40**	**17685.66**†	**17683.32**±15.58	**17685.66**•	**17683.32**±15.57	**17685.82**⊙	**17683.45**±15.55
$B_{Mm\|Q}$	**0.00297**	**0.0558**	**0.000120**	**9.19·10^{-5}** ±0.000133	**0.000120**	**9.16·10^{-5}** ±0.000133	**4.86·10^{-5}**	**4.55·10^{-5}** ±0.000109
$B_{m1\|Q}$	0.00297	0.0558	0.000120	8.52·10^{-5} ±0.000138	0.000120	8.52·10^{-5} ±0.000138	4.86·10^{-5}	4.28·10^{-5} ±0.000110
$B_{m2\|Q}$	0.00178	0.0335	2.91·10^{-7}	0	2.91·10^{-7}	0	1.78·10^{-7}	0
$B_{m3\|Q}$	5.69·10^{-5}	0.00143	8.46·10^{-6}	1.03·10^{-5} ±2.40·10^{-6}	7.99·10^{-6}	9.89·10^{-6} ±2.41·10^{-6}	2.27·10^{-6}	3.28·10^{-6} ±1.41·10^{-6}
$B_{M1\|Q}$	0.0139	0.327	0.00213	0.00392±0.00708	0.00213	0.00392±0.00708	0.000910	0.00273±0.00702
$B_{M2\|Q}$	0.00753	0.205	1.69·10^{-6}	0	1.69·10^{-6}	0	9.58·10^{-7}	0
$B_{M3\|Q}$	0.000271	0.00906	4.95·10^{-5}	8.06·10^{-5} ±8.38·10^{-6}	4.87·10^{-5}	8.04·10^{-5} ±8.65·10^{-6}	1.61·10^{-5}	3.56·10^{-5} ±9.78·10^{-6}
W_B	*5243.12*	*5119.13*	*5295.37*	*5295.90±13.14*	*5295.37*	*5295.90±13.14*	*5295.76*	*5296.16±13.30*
				$\alpha = 1.0$				
W_Q	**16024.58**×	**15998.35**	**16028.14**‡	**16077.61**±15.03	**16028.14**□	**16077.61**±15.03	**16028.14**⊗	**16077.61**±15.03
$B_{Mm\|Q}$	**2.47·10^{-5}**	**0.00678**	**2.19·10^{-8}**	**0**	**2.19·10^{-8}**	**0**	**1.78·10^{-8}**	**0**
$B_{m1\|Q}$	2.47·10^{-5}	0.00678	2.19·10^{-8}	0	2.19·10^{-8}	0	1.78·10^{-8}	0
$B_{m2\|Q}$	1.25·10^{-5}	0.00416	<1·10^{-10}	0	<1·10^{-10}	0	<1·10^{-10}	0
$B_{m3\|Q}$	3.76·10^{-7}	0.000153	1.7·10^{-9}	0	1.6·10^{-9}	0	1.6·10^{-9}	0
$B_{M1\|Q}$	0.000100	0.0530	4.76·10^{-7}	0	4.76·10^{-7}	0	4.76·10^{-7}	0
$B_{M2\|Q}$	7.07·10^{-5}	0.0298	<1·10^{-10}	0	<1·10^{-10}	0	<1·10^{-10}	0
$B_{M3\|Q}$	2.25·10^{-6}	0.00113	1.10·10^{-8}	0	1.10·10^{-8}	0	1.10·10^{-8}	0
W_B	*3314.37*	*3341.90*	*3355.88*	*3350.97±24.92*	*3355.88*	*3350.97±24.92*	*3355.88*	*3350.97±24.92*

MCF-lex-W method solution: ◁) 94.89%; ▷) 99.91%; ×) 99.73% of W_Q^{ideal} (the ideal revenue extracted from the data in [6]);
HMOR-S2: *) 99.37%; †) 100%; ‡) 99.75% of W_Q^{ideal}; HMOR-S2$_{PAS}$(i): ◇) 99.39%; •) 100%; □) 99.75% of W_Q^{ideal}; HMOR-S2$_{PAS}$(f):
⋆) 99.56%; ⊙) 100%; ⊗) 99.75% of W_Q^{ideal}.

The analytical results displayed in Tables 4 and 5 were obtained in approximately 47 s (on average) in a Linux environment on a Pentium 4 processor with 3 GHz CPU and 1 GB of RAM. In the tables, the values obtained for W_Q, $B_{Mm\|Q}$ and W_B are highlighted, as they are the most interesting o.f. (from a traffic engineering perspective) in the two priority levels.

A comparison of the results obtained with the MCF-lex-W approach (described in Subsection 3.3) and the heuristic proposed by the authors shows that the latter approach provides consistently better values for all the o.f. in most cases. This improvement is particularly relevant concerning the 'fairness' QoS o.f. $B_{Mm\|Q}$ as could be expected having in mind the nature of our model, which explicitly considers this parameter as an o.f. These results put also in evidence the superiority, especially concerning QoS related

performance parameters, of a model such as ours, which has not only an imbedded stochastic representation of the traffic flows, but also a consistent (albeit approximate) and complete representation of the interactions among all traffic flows of all types. This is naturally something that the MCF-lex-W approach cannot provide, although leading to very similar values for the QoS traffic revenue. Notice that for lower traffic loads ($\alpha = 0.5$ and $\alpha = 1.0$), the values of W_Q and W_B are very similar in both methods. This can be explained by the fact that in these situations, corresponding to low and extremely low blocking probabilities the effects of the stochasticity of the traffic are attenuated or even negligible, as indeed reflected by the values of the blocking probability related parameters in Tables 4 and 5.

Regarding the analytical results with the heuristic variants considered by the authors, Tables 4 and 5 enable two dif-

ferent comparative analysis. Since in HMOR-S2$_{PAS}$(i), the initial solution is the same as the one used in the basic heuristic HMOR-S2, the tables allow for a comparison of the final analytical results obtained with HMOR-S2 and HMOR-S2$_{PAS}$. As for the PAS(f), the initial solution has the o.f. values shown in the table under HMOR-S2 (Basis), so that a comparison of the initial and the final analytical results with HMOR-S2$_{PAS}$ can be made.

The final analytical results for the upper level o.f. are the same or show an improvement on the ones obtained with the basic heuristic, for all the values of α, for both versions of the heuristic HMOR-S2$_{PAS}$. For this reason, and also taking into account that using the archive does not lead to an increase in the execution time, the heuristic HMOR-S2$_{PAS}$ can be considered as a better approach for solving the routing problem. In particular, the (f) version (a run of the basic heuristic HMOR-S2 followed by a run of the PAS variant) provides improved results for W_Q and $B_{Mm|Q}$ for the routing problem under analysis especially for $\alpha = 0.0$, which corresponds to higher overload situations.

For $\alpha = 0.0$, the results for HMOR-S2$_{PAS}$(f) show that there was a minor improvement in the QoS flows revenue obtained with HMOR-S2, of 0.02% and 0.18% in Table 4 and 5, respectively; as for the improvement in the $B_{Mm|Q}$ value, it was significant: 17.55% and 21.95% for networks $\mathscr{G}$ and $\mathscr{H}$, respectively. For $\alpha = 0.5$ and $\alpha = 1.0$, the results are practically the same for all the versions of the heuristic. However, note that for $\alpha = 0.5$ the HMOR-S2$_{PAS}$(f) variant allowed for an improvement on the value of $B_{Mm|Q}$ in both networks.

The results presented in both tables confirm the advantages of using a Pareto archive strategy. In the situations of higher blocking ($\alpha = 0.0$), the use of this strategy leads to an improvement on the values of the first level o.f. of the routing model, especially for the blocking probability values $B_{Mm|Q}$. In the situations of lower blocking probability ($\alpha = 0.5$ and $\alpha = 1.0$), the main advantage of using the Pareto archive is the increased insensitivity to the initial solution, because for both networks the final solutions obtained with HMOR-S2$_{PAS}$(i) and HMOR-S2$_{PAS}$(f) are quite close or even the same. It should be noted that the difference between HMOR-S2$_{PAS}$(i) and HMOR-S2$_{PAS}$(f) is simply the initial solution.

3.5. Simulation Results

Simulation experiments with a static routing method using the solution provided by the heuristic were carried out. With this simulation study, the routing model results may be validated and the errors intrinsic to the analytical model, which provides the estimates for the o.f. may be evaluated.

A discrete-event stochastic simulation platform was used with the static routing model. The routing plan is the final solution obtained after one of the heuristic versions was run, and it does not change throughout the simulation, regardless of the random variations of traffic offered to the network. An initialization phase that lasts for a time $t_{warm-up}$ is fol-

lowed by a phase of data collection: information on the number of offered calls and carried calls in the network for each flow $f_s, s \in \mathscr{S}$, is gathered, until the end of the simulation. Considering this information, $B(f_s), \forall s \in \mathscr{S}$ can be estimated. Subsequently, the values of the upper and lower level o.f. related to blocking probabilities can also be estimated. The number of carried calls in the network is used to estimate the expected revenues.

In Tables 4 and 5, the analytical values and the simulation results (average value $\pm$ half length of a 95% confidence interval, computed by the independent replications method, see e.g. [19]) of each o.f. are displayed. The simulation results displayed in the table were obtained with a total simulated time $t_{total} = 48$ h and a warm-up time $t_{warm-up} = 8$ h. It took about 30 minutes of CPU time to get the results for both networks, in the computer mentioned earlier.

The analytical results and the corresponding static routing model simulation results have similar magnitude, with the analytical results slightly better than expected. The analytical and the simulation results for W_Q are close and the analytical result for that o.f. is inside the 95% confidence interval for all the heuristic versions for $\alpha = 0.0$ and $\alpha = 0.5$. For $\alpha = 1.0$, the analytical value of W_Q is actually worse than the corresponding simulation result. Notice that $\alpha = 1.0$ corresponds to a situation of lower traffic load, where in many instances all the offered calls of a certain service are actually carried. In these situations, the blocking estimate for that service is 0 and high values of the estimate of W_Q are obtained, surpassing the analytical values. Note that in lower traffic load situations ($\alpha = 0.5$ and $\alpha = 1.0$), the occurrence of blocking is a rare event. A well known result in statistics is that in these cases the uncertainty in the estimates is very high, as reflected in the very high relative half length of the calculated 95% confidence intervals of the blocking probabilities. Also for the situations of lower traffic load the simulation results for $B_{Mm|Q}$ are better than the corresponding analytical value, again because of the many instances throughout the executed simulations where the blocking estimate for a certain service is 0.

The simulation and analytic results are different mainly due to the imprecisions/inaccuracies intrinsic to the analytic/numerical resolution, in particular those associated with the simplifications of the traffic model, and the associated error propagation. In this model, the overflow traffic is treated as Poisson traffic and as a result, the analytical model is simplified and tends to underestimate the blocking probabilities in the network (and to overestimate the revenues). The errors resulting from this simplification propagate throughout the complex and lengthy numerical calculations associated with the resolution, for a great number of times, of the large systems of implicit non-linear equations used to calculate B_{ks} and $c_{ks}^{Q(B)}$. Another simplification assumed in the stochastic model for the traffic in the links is the superposition of independent Poisson flows and independent occupations of the links. However, we believe that the approximations in this model can be considered appropriate in this context for practical reasons. In fact, if

more complex models were used to represent the traffic and to calculate the blockings, the computational burden would become too heavy. Plus, these errors do not compromise the inequality relations between the o.f. values, the comparison of which is at the core of the multiobjective routing optimization method. In fact, when the results obtained with the basic heuristic HMOR-S2 and with HMOR-S2$_{PAS}$ are compared, we observe a coherence in the analytical and simulation results, in the sense that whenever the analytical value of an o.f. is better for the (f) version than for the (i) version, the same tends to happen with the average values obtained with the static routing model simulation.

4. Conclusions and Further Work

This work began with a revision of a hierarchical bi-level multiobjective routing model in MPLS networks considering alternative routing, two classes of service (with different priorities in the optimization model) and different types of traffic flows in each class. The resolution of this very complex routing optimization model was performed by a heuristic, HMOR-S2$_{PAS}$, which was also reviewed. This procedure maintains the resolution framework of a previous heuristic, HMOR-S2, but introduces and treats in a special manner an archive of possible good solutions found throughout the execution of the heuristic.

The heuristic approaches HMOR-S2 and HMOR-S2$_{PAS}$ were applied to two new test networks, $\mathscr{G}$ and $\mathscr{H}$, obtained by a transformation of an original network in [6]. Various traffic matrices were considered, so as to include in the study different situations of higher and lower traffic load.

The analytical results for the different o.f. obtained with both heuristic variants (without and with the Pareto archive) were compared. The values of W_Q were also compared with the approximate ideal values obtained with the traffic matrix provided by [6] and offered to networks $\mathscr{G}$ and $\mathscr{H}$.

Furthermore, a comparison of the results obtained with the proposed heuristic HMOR-S2$_{PAS}$ with results from a routing method based on a MCF approach, with lexicographic optimization and the possibility of traffic splitting, similar to the one in [7] was carried out. The results show that the HMOR-S2$_{PAS}$ method provides consistently better values for all the o.f. in most cases. In particular, the results for the 'fairness' QoS o.f. $B_{Mm|Q}$ are significantly better with the proposed heuristic (where this parameter is explicitly considered as an o.f.).

Concerning QoS related performance parameters, we may conclude that the stochastic representation of the traffic flows and the complete representation of the interactions among all traffic flows of all types in our model allow for better results. Nevertheless, notice that the values of W_Q and W_B are very similar in both methods for lower traffic loads ($\alpha = 0.5$ and $\alpha = 1.0$), due to the attenuated effects of the stochasticity of the traffic in these situations, corresponding to low and extremely low blocking probabilities.

The results show that the heuristic with an archive of non-dominated solutions is always advantageous, both when the blocking is higher (in this situation HMOR-S2$_{PAS}$ tends to provide improved results for the routing problem) and lower (in this situation HMOR-S2$_{PAS}$ tends to give an increased insensitivity to the initial solution).

A more exact evaluation of the results of the heuristic was accomplished with a discrete-event simulation platform. In most cases, the analytical results and the static routing model simulation results have similar magnitude. The differences between them are due to inaccuracies intrinsic to the analytic/numerical resolution, but which have not any influence in the final routing solutions.

We conclude that the results obtained with analytic and stochastic discrete-event simulation models confirm the effectiveness of the HMOR-S2$_{PAS}$ approach to route calculation and selection in multiservice networks.

An important remark is that the PAS variant is not more complex than the basic heuristic. Nevertheless, the computational burden of either resolution approach is still heavy. This is the major limitation of this type of routing method and, as so, its potential practical application is currently restrained to networks with a limited number of nodes, such as the core and intermediate (metro-core) level networks of low dimension.

Further simplifications and improvements in the heuristic resolution approaches will be the focus of future work. The extension of the model to broader routing principles (such as probabilistic load sharing or traffic splitting) and an adaptation of the model, so that it can be applied to test networks based on actual MPLS networks are also possible subjects of future work.

Acknowledgements

References

[1] A. P. Wierzbicki, "Telecommunications, multiple criteria analysis and knowledge theory", *J. Telecommun. Inform. Technol.*, no. 3, pp. 3–13, 2005.

[2] A. P. Wierzbicki and W. Burakowski, "A conceptual framework for multiple-criteria routing in QoS IP networks", *Int. Trans. Oper. Res.*, vol. 18, no. 3, pp. 377–399, 2011.

[3] J. Craveirinha, R. Girão-Silva, and J. Clímaco, "A meta-model for multiobjective routing in MPLS networks", *Central Eur. J. Oper. Res.*, vol. 16, no. 1, pp. 79–105, 2008.

[4] R. E. Steuer, *Multiple Criteria Optimization: Theory, Computation and Application*. Probability and Mathematical Statistics, Wiley, 1986.

[5] R. Girão-Silva, J. Craveirinha, and J. Clímaco, "Hierarchical multiobjective routing model in Multiprotocol Label Switching networks with two service classes – a Pareto archive strategy", *Engineering Optimization*, vol. 44, no. 5, pp. 613–635, 2012.

[6] S. C. Erbas and C. Erbas, "A multiobjective off-line routing model for MPLS networks", in *Proc. 18th Int. Teletraffic Congr. ITC-18*, J. Charzinski, R. Lehnert, and P. Tran-Gia, Eds., Berlin, Germany, 2003, Elsevier, Amsterdam, pp. 471–480.

[7] D. Mitra and K. G. Ramakrishnan, "Techniques for traffic engineering of multiservice, multipriority networks", *Bell Labs Technical J.*, vol. 6, no. 1, pp. 139–151, 2001.

[8] Q. Ma and P. Steenkiste, "Supporting dynamic inter-class resource sharing: A multi-class QoS routing algorithm", in *Proc. IEEE Infocom'99*, New York, USA, 1999, pp. 649–660.

[9] H. Kochkar, T. Ikenaga, and Y. Oie, "QoS routing algorithm based on multiclasses traffic load", in *Proc. IEEE Global Telecommun. Conf. GLOBECOM 2001*, San Antonio, TX, USA, 2001, pp. 2193–2198.

[10] D. Awduche, A. Chiu, A. Elwalid, I. Widjaja, and X. Xiao, "Overview and principles of Internet traffic engineering", RFC 3272, Network Working Group, May 2002.

[11] F. Kelly, "Notes on effective bandwidths", in *Stochastic Networks: Theory and Applications*, F. P. Kelly, S. Zachary, and I. Ziedins, Eds., vol. 4 of *Royal Statistical Society Lecture Notes Series*, Oxford University Press, 1996, pp. 141–168.

[12] D. Mitra, J. A. Morrison, and K. G. Ramakrishnan, "Optimization and design of network routing using refined asymptotic approximations", *Performance Evaluation*, vol. 36-37, pp. 267–288, 1999.

[13] J. S. Kaufman, "Blocking in a shared resource environment", *IEEE Trans. Communi.*, vol. COM-29, no. 10, pp. 1474–1481, 1981.

[14] J. W. Roberts, "Teletraffic models for the Telecom 1 integrated services network", in *Proce. 10th Int. Teletraffic Congr.*, Montreal, Canada, 1983.

[15] D. Mitra and J. A. Morrison, "Erlang capacity and uniform approximations for shared unbuffered resources", *IEEE/ACM Trans. Netw.*, vol. 2, no. 6, pp. 558–570, 1994.

[16] H. M. El Sayed, M. S. Mahmoud, A. Y. Bilal, and J. Bernussou, "Adaptive alternate-routing in telephone networks: Optimal and equilibrium solutions" *Information and Decision Technologies*, vol. 14, no. 1, pp. 65–74, 1988.

[17] J. Craveirinha, R. Girão-Silva, J. Clímaco, and L. Martins, "A hierarchical multiobjective routing model for MPLS networks with two service classes", in *Revised Selected Papers of the 23rd IFIP TC7 Conf. Sys. Model. Optimiz., Cracow, Poland, July 23-27, 2007*, A. Korytowski, K. Malanowski, W. Mitkowski, and M. Szymkat, Eds., vol. 312 of *IFIP Advances in Information and Communication Technology*, Springer, 2009, pp. 196–219.

[18] R. Girão-Silva, J. Craveirinha, and J. Clímaco, "Hierarchical multiobjective routing in Multiprotocol Label Switching networks with two service classes – A heuristic solution", *Int. Trans. Oper. Res.*, vol. 16, no. 3, pp. 275–305, 2009.

[19] A. M. Law and W. D. Kelton, *Simulation Modeling and Analysis*. Industrial Engineering and Management Science, 2nd edition. McGraw-Hill, 1991.

Rita Girão-Silva graduated in electrical engineering (telecommunications) in 1999 and received a Ph.D. diploma in electrical engineering (telecommunications and electronics) in 2009, both at the University of Coimbra, Portugal. She is an Assistant Professor at the Department of Electrical and Computer Engineering, University of Coimbra, and a researcher at INESC-Coimbra. Her research areas include routing models in telecommunications networks and multiobjective optimization.
E-mail: rita@deec.uc.pt
Department of Electrical Engineering Science
and Computers
University of Coimbra
Pólo II, Pinhal de Marrocos
P-3030-290 Coimbra, Portugal

Institute of Computers and Systems Engineering
of Coimbra (INESC-Coimbra)
Rua Antero de Quental, 199
P-3000-033 Coimbra, Portugal

José Craveirinha is Full Professor in telecommunications at the Department of Electrical Engineering and Computers of the Faculty of Sciences and Technology of the University of Coimbra, Portugal, since 1997. He obtained the following degrees: undergraduate diploma in electrical engineering science (E.E.S.)-telecommunications and electronics at IST, Lisbon Technical University (1975), M.Sc. (1981) and Ph.D. in E.E.S. at the University of Essex (UK) (1984) and Doct. of Science (Agregado) in E.E.S.-telecommunications at the University of Coimbra (1996). Previous positions were: Associate Professor and Assistant Professor at the FCTUC, Coimbra University, telecommunication R&D engineer (at the CET-Portugal Telecom). He coordinated a research group in teletraffic engineering and network planning at the INESC-Coimbra R&D Institute since 1986 and was Director of this Institute during 1994–1999. He is author/co-author of more than 100 scientific and technical publications in teletraffic modeling, reliability analysis, planning and optimization of telecommunication networks. His main present interests are in multicriteria routing and reliability analysis models and algorithms for optical and multiservice-IP/MPLS networks.
E-mail: jcrav@deec.uc.pt
Department of Electrical Engineering and Computers
University of Coimbra
Pinhal de Marrocos
3030-290 Coimbra, Portugal

Institute of Computers and Systems Engineering
of Coimbra (INESC-Coimbra)
Rua Antero de Quental, 199
P-3000-033 Coimbra, Portugal

João Clímaco is Full Professor at the Faculty of Economics of the University of Coimbra and President of the Scientific Committee of the INESC-Coimbra. He obtained the M.Sc. degree in Control Systems at the Imperial College of Science and Technology, University of London (1978); the "Diploma of Membership of the Imperial College of Science and Technology" (1978); the Ph.D. in Optimization and Systems Theory, Electrical Engineering Department, University of Coimbra (1982) and the title of "Agregação" at the University of Coimbra (1989). He served in the past as the Vice-President of ALIO – Latin-Iberoamerican Operational Research Societies Association and Vice-President of the Portuguese OR Society. He belongs to the editorial board of the following scientific journals: Investigação Operacional (Journal of the Portuguese OR Society), Journal of Group Decision and Negotiation, International Transactions in Operational Research (ITOR), ENGEVISTA and Rio's International Journal on Sciences of Industrial and Systems Engineering and Management. He is also member of the editorial board of the University of Coimbra Press. His current interests of research include multicriteria decision aiding, multiobjective mathematical programming, location analysis and telecommunication network planning and management.
E-mail: jclimaco@inescc.pt
Institute of Computers and Systems Engineering of Coimbra (INESC-Coimbra)
University of Coimbra
Rua Antero de Quental, 199
P-3000-033 Coimbra, Portugal