Instituto de Engenharia de Sistemas e Computadores de Coimbra
INESC - Coimbra

Rita Girão-Silva, José Craveirinha, João Clímaco

Stochastic Hierarchical Multiobjective Routing Model in MPLS Networks with Two Service Classes – Report on a robustness analysis experiment

9/2010 July 2010

Instituto de Engenharia de Sistemas e Computadores de Coimbra
INESC - Coimbra
Rua Antero de Quental, 199; 3000-033 Coimbra; Portugal
`www.inescc.pt`

Stochastic Hierarchical Multiobjective Routing Model in MPLS Networks with Two Service Classes - Report on a robustness analysis experiment[*][‡]

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

[a] Department of Electrical Engineering Science and Computers of the University of Coimbra
Pólo II, Pinhal de Marrocos; P-3030-290 Coimbra; Portugal
[b] Faculty of Economics of the University of Coimbra
Av. Dias da Silva, 165; P-3004-512 Coimbra; Portugal
[c] Institute of Computers and Systems Engineering of Coimbra (INESC-Coimbra)
R. Antero de Quental, 199; P-3000-033 Coimbra; Portugal

Tel.: +351 239 796252; Fax: +351 239 796247
`{rita,jcrav}@deec.uc.pt; jclimaco@fe.uc.pt`

Abstract

The report begins by reviewing a two-level hierarchical multicriteria routing model for MPLS networks with two service classes (Quality of Service and Best Effort services) and alternative routing, previously proposed by the authors. Some key issues raised by its great complexity are discussed, as well as the major factors that constitute the sources of imprecision, inaccuracy and uncertainty of the model and the way in which they are dealt with in the adopted resolution approach. Analytic and stochastic simulation experiments, for a benchmarking case study, are performed in order to evaluate the inaccuracies intrinsic to the analytic/numerical resolution procedures. Furthermore, various experiments are executed for evaluating the effects in the model results of the uncertainty associated with the estimates of the mean of the stochastic traffic flows, in a dynamic version of the routing method. In this form, a robustness analysis focused on key robustness aspects of the model was carried out.

Keywords Routing models, Multiobjective optimisation, Telecommunication networks, Simulation, Sources of uncertainty.

1 Introduction and Motivation

In modern multiservice networks, multiple and heterogeneous QoS (Quality of Service) routing requirements have to be taken into account. Therefore, the routing models designed to calculate and select one (or more) sequences of network resources (routes) have to satisfy certain QoS constraints and seek the optimisation of route related objectives.

---

[*] A shorter version was presented at the 25th Mini-EURO Conference on Uncertainty and Robustness in Planning and Decision Making (URPDM 2010), Coimbra, Portugal, Apr. 15-17 2010.

[‡] Work finantially supported by programme COMPETE of the EC Community Support Framework III and cosponsored by the EC fund FEDER and national funds (FCT).

The formulation of important routing problems in these types of networks as multiple objective optimisation problems is potentially advantageous, as these multiple objective formulations allow the trade-offs among distinct performance metrics and other network cost function(s) to be pursued in a consistent manner.

QoS issues have become increasingly relevant in the new technological platforms of multiservice networks, triggering an interest in the application of multicriteria approaches to network design and routing models in communication networks, as analysed in the reviews in [2, 4]. The technical capabilities of MPLS (Multiprotocol Label Switching) networks allow for the implementation of multiple connection-oriented services with QoS requirements. Advanced QoS-based routing mechanisms are employed in the MPLS environment and "explicit routes" (i.e. routes completely determined at the originating node) for each traffic flow of a given service type may be devised. The authors have discussed key methodological and modelling issues associated with multicriteria routing in MPLS networks in [5].

A meta-model for hierarchical multiobjective network-wide routing in MPLS networks has also been presented in [5]. In this optimisation approach, two classes of services (and different types of traffic flows in each class), are considered, QoS and BE (Best Effort) type flows. QoS flows are regarded as first priority flows and, when accepted by the network, have a guaranteed QoS level, related to the required bandwidth. As for BE flows, they are considered in the model as second priority flows, and are routed with the best possible quality of service but not at the cost of deteriorating the QoS of the QoS traffic flows. 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 optimisation typically with two optimisation levels, including fairness objectives: the first priority objective functions (o.f.) refer to the network level objectives of QoS flows, namely the total expected revenue and the maximal value of the mean blocking of all types of QoS flows; the second priority o.f. 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 generalised Erlang model for estimating the blocking probabilities in the arcs, as the one used in [20].

In this report, the main features of the routing model and the foundations of the resolution procedure in [12], are reviewed. The main focus of this work is on the analysis of the factors of imprecision, inaccuracy and uncertainty (IIU, for short), see [1], of the routing model and their effects on the results of the routing method. The major factors that constitute the sources of IIU of the model and of the adopted resolution approach are analysed in detail. The imprecision factors stem mainly from the approximations inherent to the analytic traffic model (underlying the optimisation model) and the numerical errors in the calculation of blocking probabilities throughout the resolution procedures. The inaccuracies are a consequence of the extreme complexity of the multicriteria optimisation model, and are related to the interdependencies of the o.f. These interdependencies result from the fact that all the specified o.f. depend on all traffic flow patterns in the global network, which change with any alteration in a route used by any given node to node flow. This mechanism generates potential instability in the global routing solutions, as analysed in a similar, albeit simpler model, shown in [19]. As for the uncertainty issues, they are raised mainly by the stochastic nature of the offered traffic and the related estimation procedures performed throughout the stochastic simulation experiments with the proposed heuristic.

Analytical and stochastic simulation experiments with the proposed heuristic in a network case study are performed in order to evaluate the inaccuracies intrinsic to the analytic/numerical resolution procedures. Two types of discrete event stochastic simulation were performed, one with a static routing model where the network routing plan never changes, and another with a state-dependent periodic-type dynamic routing model where the network routing plan is updated as a function of the measured offered traffic in the network. Different simulation parameters have a direct influence on the robustness of the results obtained with the dynamic model simulation. In particular, the effects in the model results of the uncertainty associated with the moving average estimates of the offered traffic in the network and the influence of the routing plan update time interval are analysed, allowing for a robustness analysis focused on key robustness aspects of the model.

The major contributions of this report are the following:

1. An analysis of the major factors that constitute the sources of imprecision, inaccuracy and uncertainty of a very complex multiobjective routing model for MPLS networks with two service classes, which uses o.f. expressed in terms of stochastic measures;

2. Presentation of an analytic and stochastic simulation study enabling to evaluate the inaccuracies intrinsic to the analytic/numerical resolution procedures and their possible impact on the results of the heuristic resolution procedures of the model;

3. Development of an experimental study, using discrete-event stochastic simulation, for evaluating the effects in the model results of the uncertainty associated with the statistical estimates associated with the traffic flows, in a dynamic version of the routing method;

4. A description of the way in which the developed resolution procedures deal with the above difficult issues.

The report is organised as follows. The two-level hierarchical multiobjective alternative routing model with two service classes and the features of the considered heuristic are briefly reviewed in section 2. In section 3, an analysis of the IIU factors associated with key instances of the model and its resolution and their potential effects on the heuristic results, is provided. The results obtained with this procedure, by using analytic and discrete-event simulation experiments for a test network used in a benchmarking study, are shown and analysed in relation to the IIU factors, in the fourth section. These experiments provide some preliminary conclusions on the robustness of the method concerning some of the IIU factors. Conclusions are drawn in the last section. Finally, the report ends with an appendix with the specification of the notation used in the model.

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

2.1 The Multiobjective Routing Model

The considered model is an application of the multiobjective modelling framework (or “meta-model”) for MPLS networks proposed in [5], as previously mentioned. It is a network-wide routing optimisation approach (that is, the main o.f. depend explicitly on all traffic flows in the network) of new type, in the form of a hierarchical multiobjective optimisation model, which takes into account the nature and relations between the adopted o.f. related to the different types of traffic flows associated with different services.

Two classes of services are considered: 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 different service types of each class are represented through the sets $\mathcal{S}_Q$ (for QoS service types) and $\mathcal{S}_B$ (for BE service types). The traffic flows of each service type $s \in \mathcal{S}_Q$ or $s \in \mathcal{S}_B$ may differ in important attributes, in particular the required bandwidth.

The hierarchical multiobjective routing optimisation model considered here has two levels with several o.f. in each level. The first level includes the first priority o.f. (the total expected network revenue associated with QoS traffic flows, W_Q, and the worst average performance among QoS services, represented by the maximal average blocking probability among all QoS service types, $B_{Mm|Q}$), which are formulated at the network level for the QoS traffic and considering the combined effect of all types of traffic flows in the network. In the second level the o.f. are concerned with average performance metrics of the QoS traffic flows associated with the different types of QoS services (represented by the mean blocking probabilities for flows of type $s \in \mathcal{S}_Q$, $B_{ms|Q}$, and the maximal blocking probability $B_{Ms|Q}$, defined over all flows of type $s \in \mathcal{S}_Q$) as well as the total expected network revenue associated with BE traffic flows, W_B. These constitute the second priority o.f. At the two levels of optimisation, ‘fairness’ objectives are explicitly considered in the form of $\min - \max$ objectives. These objectives seek to make the most of the proposed multiobjective formulation. In previous formulations of routing models for these networks, such type of 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 following the same type of approach as in [20].

The considered two-level hierarchical optimisation problem for two service classes P-M2-S2 (‘**P**roblem - **M**ultiobjective with **2** optimisation hierarchical levels - with **2** **S**ervice classes’) is:

Problem P-M2-S2

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

and with $W_{Q(B)} = \sum_{s\in\mathcal{S}_{Q(B)}} A^c_s w_s$; $B_{Mm|Q} = \max_{s\in\mathcal{S}_Q}\{B_{ms}\}$; $B_{ms|Q} = \frac{1}{A^o_s} \sum_{f_s\in\mathcal{F}_s} A(f_s)B(f_s)$;

and $B_{Ms|Q} = \max_{f_s \in \mathcal{F}_s}\{B(f_s)\}$. The total traffic offered by flows of type s is A_s^o, the carried traffic for service type s is A_s^c, the mean traffic offered associated with $f_s \in \mathcal{F}_s$ is $A(f_s)$ (all in Erlang); $B(f_s)$ is the node to node blocking probability for all flows $f_s \in \mathcal{F}_s$ and w_s is the expected revenue per call of service type s.

Note that in the formulation of P-M2-S2 while W_Q is a first priority o.f. (together with $B_{Mm|Q}$), W_B will be a second level o.f. This guarantees that the routing of BE traffic, in a quasi-stationary situation, will not be made at the cost of the decrease in revenue or at the expense of an increase in the maximal blocking probability of QoS traffic flows. Nevertheless, it is important to note that while QoS and BE traffic flows are treated separately in terms of o.f. so as to take into account their different priority in the routing optimisation, the interactions among all traffic flows are fully represented in the model. This is assured by the used traffic modelling approach, underlying the optimisation model, because the traffic model used to obtain the blocking probabilities $B(f_s)$ integrates the contributions of all traffic flows which may use every link of the network. This feature is a major difference in comparison with more common routing models that have been proposed for networks with two service classes, based on some form of decomposition of the network representation, corresponding to 'virtual networks', one for each service class (e.g. in [21]).

A full description of the traffic modelling stochastic approach used in the routing model can be seen in [5]. The basic teletraffic model allows for the blocking probabilities B_{ks}, for micro-flows of service type s in link l_k, to be given in the form $B_{ks} = \mathcal{B}_s\left(\overline{d_k}, \overline{\rho_k}, C_k\right)$, where $\mathcal{B}_s$ represents the basic function (implicit in the teletraffic analytical model) that expresses the marginal blocking probabilities, B_{ks}, in terms of $\overline{d_k} = \left(d_{k1}, \cdots, d_{k|\mathcal{S}|}\right)$ (vector of equivalent effective bandwidths for all service types), $\overline{\rho_k} = \left(\rho_{k1}, \cdots, \rho_{k|\mathcal{S}|}\right)$ (vector of reduced traffic loads ρ_{ks} offered by flows of type s to l_k) and the link capacity C_k. This type of approximation (see [20]) enables the calculation of $\{B_{ks}\}$ through efficient and robust numerical algorithms, which are essential in a network-wide routing optimisation model of this type, for tractability reasons.

The decision variables $\overline{R} = \cup_{s=1}^{|\mathcal{S}|} R(s)$ 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 $R(s) = \cup_{f_s \in \mathcal{F}_s} R(f_s), s \in \mathcal{S}_Q \cup \mathcal{S}_B$ and $R(f_s) = (r^p(f_s)), p = 1, \cdots, M$ with $M = 2$ in our model. An alternative routing principle is used: for each flow f_s the first choice route $r^1(f_s)$ will be used; if it is blocked the routing method makes the connection request attempt the second choice route $r^2(f_s)$. A request will be blocked only if $r^2(f_s)$ is also blocked.

The very high complexity of the routing problem P-M2-S2 stems from two major 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. Note that all these interdependencies are defined explicitly or implicitly through the underlying traffic model. Also note that even in the simplest degenerated case (single service with single-criterion optimisation and no alternative routing) the problem is NP-complete in the strong sense, as proved in [10]. Having in mind the form of P-M2-S2, one may conclude on the great intractability of this problem. There are possible conflicts between the o.f. in P-M2-S2, because in many routing situations, the maximisation of W_Q leads to a deterioration on some $B(f_s)$, $s \in \mathcal{S}_Q$, for certain traffic flows with low intensity, and this tends to increase $B_{Ms|Q}$ and $B_{ms|Q}$, and consequently $B_{Mm|Q}$. This is a major factor to justify the interest and potential advantage in using multiobjective approaches when dealing with this type of routing methods.

2.2 The Heuristic Resolution Approach

The theoretical foundations of a specialised heuristic strategy for finding "good" compromise solutions to the very complex bi-level hierarchical multiobjective alternative routing optimisation problem, were presented in [7]. In [13], a basic heuristic approach (HMOR-S2 or **H**ierarchical **M**ulti**o**bjective **R**outing considering **2** classes of **S**ervice) devised to find "better" solutions to this problem, was proposed and applied to a test network used in a benchmarking case study, for various traffic matrices. A new variant of this resolution method, HMOR-S2$_{\mathrm{PAS}}$ (HMOR-S2 with a **P**areto **A**rchived **S**trategy), based on the introduction of a Pareto archive in the basic heuristic, inspired in one of the standard procedures used in Pareto archived evolutionary meta-heuristics [17], is proposed in [12].

The resolution (in a multicriteria analysis sense) of the routing problem P-M2-S2 is based on the recurrent calculation of solutions to a constrained bi-objective shortest path problem, formulated for every end-to-end flow f_s:

$$\text{Problem } \mathcal{P}^{(2)}_{s2} : \min_{r(f_s)\in\mathcal{D}(f_s)} \left\{ m^n(r(f_s)) = \sum_{l_k\in r(f_s)} m^n_{ks} \right\}_{n=1;2}$$

The path metrics m^n to be minimised are the marginal implied costs $m^1_{ks} = c^{Q(B)}_{ks}$ and the marginal blocking probabilities $m^2_{ks} = -\log(1 - B_{ks})$; $\mathcal{D}(f_s)$ is the set of all feasible loopless paths for flow f_s, which satisfy specific traffic engineering constraints (other than the effective bandwidth) for flows of type s. A typical constraint is a maximal number of arcs per path depending on the class and type of service s. By using this approach, the comparison of the efficiency of different candidate routes in the context of a multicriteria routing framework of this type takes into account both the loss probabilities experienced along the candidate routes and the knock-on effects upon the other routes in the network, effects associated with the acceptance of a call on that given route. Such effects can be measured exactly through the marginal implied costs for QoS(BE) traffic, $c^{Q(B)}_{ks}$, associated with the acceptance of a connection (or "call") of traffic f_s of any service type $s \in \mathcal{S}$ on a link l_k, that can be defined as the expected value of the traffic loss induced on all QoS(BE) traffic flows resulting from the capacity decrease in link l_k. By a conjecture in [7], they can be obtained by solving a system of implicit non-linear equations. For more details on this important mathematical concept in routing optimisation in loss networks, see [16, 20].

It is important to note that this auxiliary constrained bi-objective shortest path problem was used as a basis of the heuristic approach having in mind that the consideration of the metric blocking probability tends, at a network level, to minimise the maximal node-to-node blocking probabilities $B(f_s)$, while the metric implied cost tends to maximise the total average revenue W_T in a single class multiservice loss network [9, 19]. However we must draw attention to the exact meaning of the statement that the minimisation of path implied cost 'tends' to maximise W_T. This would be, in rigour, only valid if the choice of such 'optimal' path, for a given f_s, would not change in any form the network working conditions concerning all the remaining traffic flows, an assumption obviously not true, due to the interdependencies among $\{c_{ks}\}$, $\{B_{ks}\}$ and $\overline{R}$. Naturally this is yet another source of difficulty in devising a heuristic of this nature for this complex routing problem (P-M2-S2).

In the heuristic, the auxiliary constrained shortest path problem $\mathcal{P}^{(2)}_{s2}$ is solved by an algorithmic approach, MMRA-S2 (**M**odified **M**ultiobjective **R**outing **A**lgorithm for multiservice networks, considering **2** classes of **S**ervice) [7], which aims at finding a 'best' compromise path from a set of non-dominated solutions[1], according to some system of preferences. In this context, path computation and selection have to be fully automated. Therefore the system of preferences is embedded in the working of the algorithm. This is implemented by defining preference regions in the o.f. space obtained from aspiration and reservation levels (preference thresholds) defined for the two o.f. [9, 19].

Another important part of the addressed routing model is the underlying traffic model. This stochastic traffic model involves all the sub-models and associated numerical procedures, that are needed for obtaining all traffic related parameters, namely implied costs and blocking probabilities B_{ks} and $B(f_s)$, under certain simplifying assumptions. A description of the traffic modelling approach used in the routing model can be seen in [5].

In the dedicated heuristic HMOR-S2, each new solution is obtained by processing the current best solution. A basic searching strategy is to seek for routing solutions $R(s)$ for each service $s \in \mathcal{S}$, in order to achieve a better performance in terms of W_B (if $s \in \mathcal{S}_B$) or $B_{ms|Q}$ and $B_{Ms|Q}$ (if $s \in \mathcal{S}_Q$), while respecting the hierarchy of o.f. This also means that network resources are left available for traffic flows of other services so that the solutions selected at each step of the procedure may improve the first priority o.f. W_Q and $B_{Mm|Q}$. The heuristic was designed in order to seek, firstly for each QoS service and starting from the services with higher effective bandwidth (considering the numbering of $s, s = 1, \cdots, |\mathcal{S}_Q|$) and, secondly, for each BE service (also beginning by the higher bandwidth services, $s = |\mathcal{S}_Q| + 1, \cdots, |\mathcal{S}|$), solutions which dominate the current one, in terms of $B_{ms|Q}$ and $B_{Ms|Q}$ for QoS services and in terms of W_B for BE services. These solutions will only be accepted if they do not lead to the worsening of any of the network functions W_Q and $B_{Mm|Q}$.

The candidate solutions $(r^1(f_s), r^2(f_s))$ for each f_s are generated using the mentioned algorithm MMRA-S2. They are selected (or rejected) according to specific criteria, to be 'tuned' throughout the execution of the heuristic. A maximal number of arcs D_s per route for each service type s is previously defined and a feasible route set $D(f_s)$ is obtained for each f_s.

The theoretical analysis of the model, confirmed by experimentation, showed that the successive application of MMRA-S2 to every traffic flow does not lead to an effective resolution approach to the network routing problem P-M2-S2. This results from an instability phenomenon that arises in such path selection procedure, expressed by the fact that the route sets $\overline{R}$ often tend to oscillate between certain solutions some of which may lead to poor global network performance under the prescribed metrics. This instability phenomenon is associated with the complexity and interdependencies in the addressed problem P-M2-S2, namely the interdependencies between $\{c_{ks}^{Q(B)}\}$ and $\{B_{ks}\}$ and between these two sets and the current network route set $\overline{R}$.

Therefore, another core idea of the heuristic approach is the search for the subset of the path set, the elements of which should be possibly changed in the next route improvement cycle. Detailed analysis and extensive experimentation with the heuristic led to the proposal of a criterion for choosing candidate paths for possible routing improvement by increasing order of a function $\xi(f_s)$ of the current $(r^1(f_s), r^2(f_s))$ [13]. With this criterion, preference (concerning the potential value in changing the second choice route when seeking to improve W_Q or W_B) 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

---

[1] In multiobjective optimisation [23], the concept of optimal solution is replaced by the concept of non-dominated (or Pareto optimal) solution. A non-dominated solution is a feasible solution such that, in minimisation problems, it is not possible to decrease the value of an o.f. without increasing on, at least, the value of one of the other o.f.

have worse end-to-end blocking probability. Another key point tackled by the heuristic is the specification of a variable $nPaths$, which represents the number of routes with smaller values of $\xi(f_s)$ that should possibly be changed by running MMRA-S2 once again [13]. In order to do so, the effect of each candidate route on the relevant o.f. is anticipated by solving the corresponding analytical model.

A variant of this basic resolution method, HMOR-S2$_{\mathrm{PAS}}$, is proposed in [12]. The realisation that throughout the execution of the basic heuristic HMOR-S2 there were interesting solutions to the routing problem that were not further pursued due to the strict limitations imposed on the acceptance of a new solution motivated the development of a new variant that could store these possibly interesting solutions in an archive and later go through them in order to try and find the "best" possible solution to the problem in hand. The numerical complexity of HMOR-S2 and HMOR-S2$_{\mathrm{PAS}}$ is the same.

The management of the archive in terms of addition and removal of solutions from the archive is governed by a set of rules thoroughly described in [12]. At the end of the algorithm (after the outer cycle), the solutions stored in the archive are analysed. By employing the concept of preference thresholds, the priority regions in the bidimensional o.f. space are re-evaluated so as to select the final solution of the algorithm.

The main features of this approach based on the definition of preference thresholds in the o.f. space are: i) the representation of QoS requirements through requested (or aspirational) and acceptable (or reservation) thresholds for each network function W_Q and $B_{Mm|Q}$; ii) the consideration of this type of thresholds defines priority regions in the bidimensional o.f. space in which non-dominated solutions may be searched for. As an example of the definition of priority regions in the bidimensional o.f. space of the solutions in the archive, see figure 1.

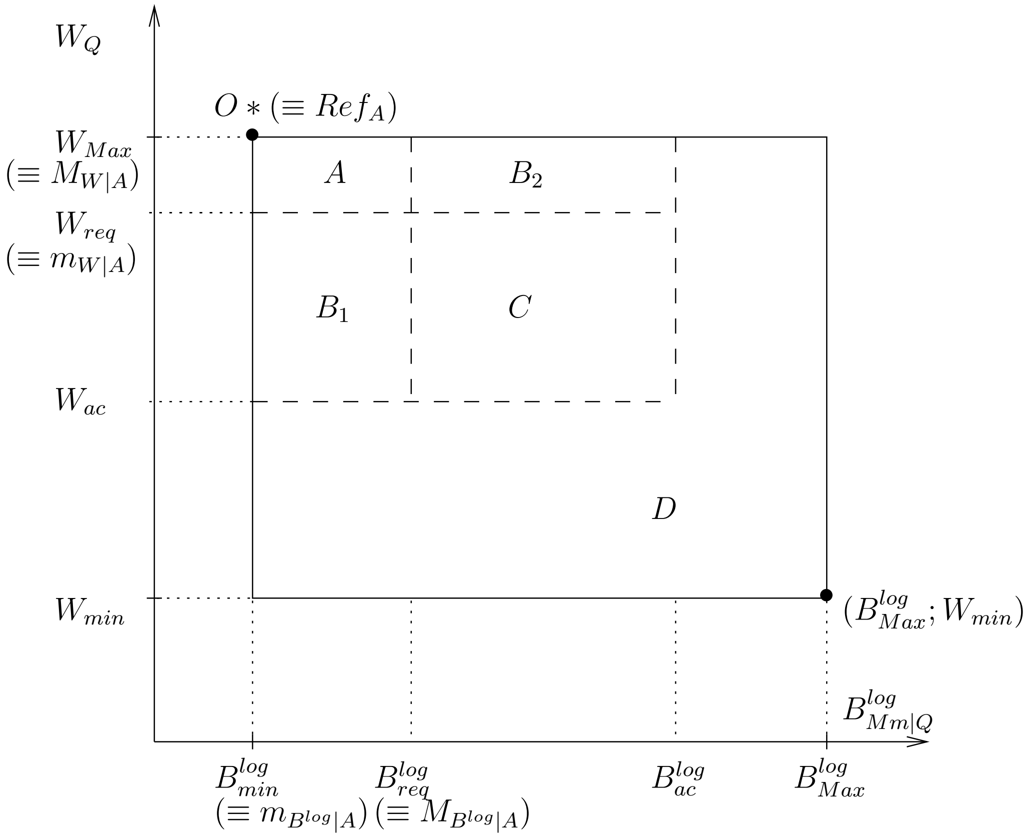

Figure 1: QoS requirements used to define priority regions in the bidimensional o.f. space

The ideal optimum is represented by $O*$ and is obtained when both first level o.f. W_Q and $B_{Mm|Q}$ are optimised. The point $(B^{log}_{Max}; W_{min})$ is the Nadir point, that is, the point with the worst values for each o.f. when the other o.f. is optimised. The first priority region A is defined by the points for which the requested levels are satisfied for both o.f. The second priority regions B_1 and B_2 are those for which only one of the requested values is satisfied and an acceptable value is guaranteed for the other metric. Although regions B_1 and B_2 have the same priority in theory, in practice the maximisation of the QoS traffic revenue is considered more important than the minimisation of $B_{Mm|Q}$. Therefore, B_2 will be considered preferable to B_1. A third priority region C, where only acceptable values are guaranteed for both metrics, is defined. Beyond the acceptable values, lies the least priority region D, that defines worst values W_{min} and B^{log}_{Max} for both metrics. The preference thresholds used to define the priority regions are calculated in a fully automated manner (see the expressions for these thresholds in Appendix A).

The approach chosen to select the "best" solution in the best possible priority region is based on the calculation of a Chebyshev distance to a reference point, as described in [3]. Note that this operation is more time-consuming than simply choosing the first solution found in the best possible priority region. However, this operation is methodologically more correct and as it is performed only once, the amount of time it takes is not of primary importance.

As described in [24], reference point type approaches minimise the distance of the solutions in the o.f. space to a selected point (often considered as the ideal one) according to a specific metric, which in our procedure is the weighted Chebyshev distance. Let $\mathcal{R}$ be the best possible priority region in the o.f. space where at least one solution ϱ can be found. In our approach, only this region $\mathcal{R}$ will be swept. A specific reference point is chosen in region $\mathcal{R}$ as the ideal point in that region, $\left(\mathcal{C}^*_{1|\mathcal{R}}; \mathcal{C}^*_{2|\mathcal{R}}\right)$. The two metrics in the region are related to the upper level o.f. of the problem P-M2-S2, $B^{log}_{Mm|Q}$ (which has to be minimised) and W_Q (which has to be maximised). That is, the ideal point in each rectangular region is the top left corner of that region. See the example in figure 1 where the reference point for region A (Ref_A) is displayed. For a non-rectangular region such as D the ideal point of the whole o.f. space $O*$ is the reference point.

Another set of parameters that must be defined is the minimum $m_{i|\mathcal{R}}$ and maximum $M_{i|\mathcal{R}}$ of each metric i for region $\mathcal{R}$. See the example in figure 1 where the minimum and maximum for both metrics in region A, are displayed. Notice that the reference point for each region can be written as $\left(\mathcal{C}^*_{1|\mathcal{R}}; \mathcal{C}^*_{2|\mathcal{R}}\right) = \left(m_{B^{log}|\mathcal{R}}; M_{W|\mathcal{R}}\right)$, because it is the point in the region where the $B^{log}_{Mm|Q}$ metric is minimised and the W_Q metric is maximised.

The problem to be solved is $\min_{\varrho\in\mathcal{R}} \max_{i=1,2} \left\{ w_{i|\mathcal{R}} \left| \mathcal{C}_i(\varrho) - \mathcal{C}^*_{i|\mathcal{R}} \right| \right\}$ where the metrics for solution ϱ are $\mathcal{C}_1(\varrho) = B^{log}_{Mm|Q}(\varrho)$ and $\mathcal{C}_2(\varrho) = W_Q(\varrho)$. The weights in the weighted Chebyshev distance are $w_{i|\mathcal{R}} = \frac{1}{M_{i|\mathcal{R}} - m_{i|\mathcal{R}}}$, which allow the Chebyshev metrics $\left\{ w_{i|\mathcal{R}} \left| \mathcal{C}_i(\varrho) - \mathcal{C}^*_{i|\mathcal{R}} \right| \right\}$ to be dimension free and proportional to the size of the rectangular region. These weights also account for the difference in range of the parameters $B^{log}_{Mm|Q}$ and W_Q.

3 Dealing with Imprecision, Inaccuracy and Uncertainty in the Model

The high complexity of the routing problem P-M2-S2 stems from the combinatorial nature of the global routing multiobjective optimisation problem and is related to two major factors specific to the described model: 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. The o.f. depend on all traffic flow patterns in the global network, and they may change significantly with any alteration in any route choice for any given node to node flow. The potential instability in the global routing solutions generated by this routing mechanism in a similar yet simpler model was analysed in [19]. Note that all these interdependencies are defined explicitly or implicitly through the underlying traffic model.

Also note that even in the simplest degenerated case (single service with single-criterion optimisation and no alternative routing) the problem is NP-complete in the strong sense, as proved in [10]. This great complexity leads to inaccurate results, as the solutions of the routing problem are inherently approximate, because the resolution method does not calculate all the non-dominated solutions and even for those selected solutions which are computed there is no certainty that they are not potentially dominated by other solutions. Concerning the possibility of not detecting the condition of certain weakly dominated solutions this may be explained by small variations in the values of some o.f. for certain solution(s) close to the current one (for example by changing a single route for a given flow), solution(s) which, in some cases, may not be detected by the heuristic. It must be remarked that this occurrence is rare but in principle may arise in any of the variants of the heuristic [13], including those which incorporate meta-heuristic techniques, namely simulated annealing or tabu search, as shown in [11]. The mechanisms of the heuristic resolution in its present version (HMOR-S2_{PAS}), as described in the previous sub-section, are devised to deal with this issue in order to minimise the impact of these inaccuracies in the quality of the obtained compromise solutions.

Several simplifications and approximations were assumed in the stochastic model for the traffic in the links, namely a superposition of independent Poisson flows and independent occupations of the links, leading inevitably to intrinsic imprecisions in the values of the traffic parameters which are reflected in the calculation of the o.f. values. Still, if a more accurate and realistic representation of the traffic flows was used, better estimates of the blocking probabilities would be achieved. Nonetheless, the approximations in our model can be considered appropriate and, above all, absolutely necessary in this context, for practical reasons. In fact, if more exact stochastic models were used to represent the traffic and to calculate the blockings in overflow conditions, the computational burden would be too heavy since the analytical model has to be numerically solved many times during the execution of the heuristic and the routing method would become intractable. Furthermore there are also numerical errors, associated with the resolution of the system of equations of the traffic model which propagate throughout the resolution procedure, as the resolution of the traffic model has to be performed many times (in the calculations of marginal implied costs and blocking probabilities of all the flows). To minimise the latter imprecision effects, some ‘robust’ and well tested numerical algorithms (namely the Kaufman/Roberts algorithm [14, 22] and fixed-point iterators [15]) are used to estimate the blocking probabilities in the system. Moreover, these two types of errors do not compromise the inequality relations between the o.f. values, as the aim of the routing optimisation procedure is just the comparison of routing solutions in terms of the values of the o.f. That is, the focus is on the relative value of the results of the traffic model rather than on the absolute accuracy of such values. Also, small differences between o.f. values can be disregarded when comparing solutions.

Further imprecision effects stem from an instability phenomenon which may potentially arise in the path selection procedure. In fact, the route sets $\overline{R}$, if obtained by successive application of MMRA-S2 to every traffic flow, often tend to oscillate between certain solutions some of which may lead to poor global network performance under the prescribed metrics, thus leading to uncertainty in the results. The experimentations confirmed that the successive application of MMRA-S2 to every traffic flow does not lead to an effective resolution approach to the network routing problem P-M2-S2. For dealing with this issue in a successful manner, detailed analysis and extensive experimentation with the heuristic led to the proposal of a criterion for choosing candidate paths for possible routing improvement by increasing order of a function $\xi(f_s)$ of the current routes of a flow f_s, giving preference (concerning the potential value in changing the routes when seeking to improve the QoS(BE) traffic revenue) to the flows for which the first route has a low implied cost and the second route has a high implied cost, or to the flows which currently have worse end-to-end blocking probability.

Finally, the stochastic nature of the traffic offered to the network leads to another form of uncertainty in the results. In the discrete event stochastic simulation experiments performed with a static routing model, a measurement of the degree of uncertainty of the o.f. values could be obtained by applying a classical statistical procedure (batch means with independent replications). As for the state-dependent periodic-type dynamic routing model, the traffic flows means are periodically updated via a statistical estimate (first-order moving average) based on real-time measurements, dependent on a parameter b (fixed a priori), as explained in sub-section 4.3, which is another source of statistical uncertainty. This required a sensitivity/robustness analysis, for the evaluation of the influence of this parameter and the routing plan update time interval on the final global routing solution.

A summary of the sources of IIU and the way these aspects are dealt with in our resolution approach are presented in table 1.

In the next section, the analytical and simulation results obtained with the HMOR-S2$_{\mathrm{PAS}}$ heuristic in a specific network case study based on the one in [21], are presented. The analytical results are obtained by a single run of the resolution heuristic and may be compared with the results presented in [21]. As for the simulation results, they allow for a more realistic assessment of the results of the heuristic having in mind the combined effects of the analysed IIU factors. Two types of simulation were performed, one corresponding to a static routing model where the routing plan calculated by the heuristic is never changed regardless of the random variations in offered traffic throughout the simulation, for a given matrix of mean traffic offered in statistical equilibrium. The other corresponds to a periodic type state-dependent dynamic routing model, where the routing plans are updated periodically as a function of real-time traffic measurements, by using the heuristic HMOR-S2 repeatedly. Dynamic routing in a telecommunications network is a well known routing principle where the most recent information on the network conditions is taken into account in order to find appropriate paths for the connection requests in the network. This is especially important when there are significant fluctuations of the offered traffic in various parts of the network, in particular as a result of overload or network failures.

Table 1: Sources of IIU and ways to deal with them

Sources of IIU	Effect	Dealing with IIU
High complexity of the routing problem P-M2-S2	Inaccuracy (the solutions are inherently approximate)	Different mechanisms of the heuristic resolution in its present version (HMOR-S2$_{\mathrm{PAS}}$)
Simplifications and approximations assumed in the stochastic model for the traffic in the links, leading to an approximate model, unavoidable for computational tractability reasons	Imprecision	The focus was on the relative value of the results of the traffic model; small differences between o.f. values can be disregarded when comparing solutions (this was achieved by using adequate numerical traffic calculation procedures)
Instability phenomenon potentially arising in the path selection procedure if the MMRA-S2 is applied successively to all the end-to-end flows of each service type	Imprecision and uncertainty	Criterion for choosing candidate paths for possible routing improvement, embedded in the main heuristic 'optimisation' cycle
Numerical errors in the calculations of marginal implied costs and blocking probabilities of all the flows, propagating throughout the resolution procedure	Imprecision	'Robust' and well tested numerical algorithms (namely the Kaufman/Roberts algorithm and fixed-point iterators)
Stochastic nature of the traffic offered to the network	Uncertainty	Periodical update of traffic flows means via a statistical estimate (first-order moving average) based on real-time measurements; sensitivity/robustness analysis

4 Experimental Results

In this section, the analytical and simulation results obtained with the HMOR-S2_{PAS} heuristic in a network case study analogous to the one in [21] will be presented.

In [21] a model for traffic routing optimisation and admission control in multiservice networks supporting traffic with different QoS requirements, was proposed. This model for MPLS networks with two service classes uses a lexicographic optimisation formulation, including admission control for BE traffic, based on a deterministic MCF (Multicommodity Flow) model, with the expected revenues associated with QoS and BE traffic as o.f. It will be used as a benchmarking study for the present work concerning upper bounds $W_Q^{\max}$ for the optimal value of the QoS traffic revenue W_Q. For a brief summary of this application model, see [13].

4.1 Application of the Model to a Network Case Study

The routing model in [21] and our routing model were applied to the test network $\mathcal{M}$ depicted in figure 2. It has $|\mathcal{N}| = 8$ nodes, with 10 pairs of nodes linked by a direct arc and a total of $|\mathcal{L}| = 20$ unidirectional arcs. The bandwidth of each arc C'_k [Mbps] is shown in figure 2. The number of channels C_k is $C_k = \left\lceil \frac{C'_k}{u_0} \right\rceil$, with basic unit capacity $u_0 = 16$ kbps. There are $|\mathcal{S}| = 4$ service types with the features displayed in table 2. The values of the required effective bandwidths $d_s = \frac{d'_s}{u_0}$ [channels] $\forall s \in \mathcal{S}$ are also in the table (where d'_s is the required bandwidth in kbps). The expected revenue for a call of type s is assumed to be $w_s = d_s, \forall s \in \mathcal{S}$. The average duration of a type s call is h_s and D_s represents the maximum number of arcs for a type s call.

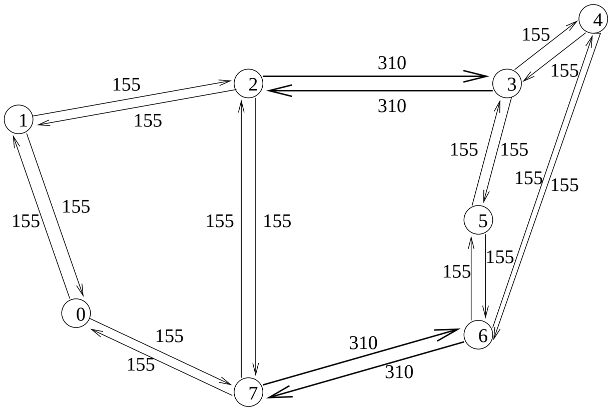

Figure 2: Test network $\mathcal{M}$ [21], with the indication of the bandwidth of each arc C'_k, in Mbps

A base matrix $T = [T_{ij}]$ with offered total bandwidth values from node i to node j [Mbps] is provided in [21]. From these data all the parameters needed by our traffic model can be obtained as shown in [6].

In this application example, results for the QoS flows revenue W_Q are presented for three values of a compensation parameter α: $\alpha = 0.0$ corresponds to a deterministic situation; $\alpha = 0.5$ 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 $\alpha = 1.0$ is used for traffic flows with higher 'variability'.

Table 2: Service features on the test network $\mathcal{M}$

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

For further details on the application of this traffic model to the network case study under analysis, see [6].

4.2 Analytical Results

In the analytical study, the HMOR-S2 heuristic (without the Pareto archive) was run once, followed by a run of the HMOR-S2$_{\text{PAS}}$ heuristic.

The initial solution is typical of Internet routing conventional algorithms: only one path for each flow (i.e. without an alternative path) is considered; the initial solution is the same for all the services $s \in \mathcal{S}$ and the paths 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 routing plan obtained at the end of the HMOR-S2 runs for each specific α is the initial solution of the HMOR-S2$_{\text{PAS}}$. This heuristic is run only once. For the archived routing plans obtained at the end of this single run, values for all the o.f. are computed and the "best" possible solution in the best possible preference region is chosen to be the final solution of the algorithm, using a reference point-based procedure as the solution selection mechanism.

4.3 Simulation Results

After the analytical experiences were performed, simulation experiences, with static and dynamic routing methods using the solution provided by the heuristic, were also carried out. This simulation study enables the validation of the routing model results and the evaluation of the errors intrinsic to the analytical model which provides the estimates for the o.f.

Initially, the discrete-event stochastic simulation was applied to a static routing model, where the routing plan is the final solution obtained after the HMOR-S2$_{\text{PAS}}$ was run. This routing plan does not change throughout the simulation regardless of the random variations of traffic offered to the network. The simulation starts with an initialisation phase that lasts for a time $t_{warm-up}$, that should be long enough to guarantee that the system state at the end of the initialisation phase is representative of the steady state behaviour of the system. After this time, information on the number of offered calls and carried calls in the network for each flow $f_s, s \in \mathcal{S}$, is gathered, until the end of the simulation. With this information, $B(f_s), s \in \mathcal{S}$ and subsequently, the values of the upper and lower level o.f. related to blocking probabilities can be estimated. The calculation of the expected revenues is based on the number of carried calls in the network.

In the periodic and state-dependent dynamic version of the routing method considered here, the network state is assessed periodically and the gathered information on that state is used to periodically choose the most appropriate paths in the network, according to the HMOR-S2 routing algorithm. The offered traffic estimate is calculated as in [8]. In the time interval $[n\tau;(n+1)\tau[$, the estimate of the average traffic offered to the network by the flow f_s is given by $\tilde{x}_n(f_s)$, a first order moving average iteration of the type $\tilde{x}_n(f_s) = (1-b)\tilde{x}_{n-1}(f_s) + b\tilde{X}_{n-1}(f_s)$, where $\tilde{X}_{n-1}(f_s)$ is an estimator of the average value of the traffic offered by f_s to the network in the previous interval $[(n-1)\tau;n\tau[$ and $b \in]0.0;1.0[$ is a compromise value between the need to obtain a quick response of the estimator to rapid fluctuations in $\tilde{X}(f_s)$ and the stability of the long-run variations and should be settled by extensive experimentation with the simulation model. The parameter τ is both the update period of the estimates of the offered traffic and the update period of the network routing plans. Note that the process of path choice should take a short time, when compared to τ.

The routing method seeks to obtain new routing plans adapted to the changing network working conditions resulting from the random fluctuations of traffic intensities. Taking into account the features and great complexity of the routing model, the choice of the "start-up" routing solution of the dynamic method is of great importance concerning its performance. The start-up solution is the final solution obtained after the HMOR-S2$_{\mathrm{PAS}}$ was run. The availability of a good estimate of the initial nominal traffic matrix is a necessary requirement of this dynamic routing method. Having in mind the periodically updated characteristics of the offered traffic, the "best" possible set of paths are chosen so as to improve the multidimensional network performance, as specified by the multiobjective routing model.

A first phase of simulation, the initialisation phase, lasts for a time $t_{warm-up} = t_0 + t_1$. In a first stage that lasts t_0, only periodical updates of the estimate of the offered traffic are performed, with period τ. After that time t_0, the offered traffic estimates are assumed to be representative of a steady state behaviour. Afterwards, during a time t_1, the estimate of the offered traffic is still performed with a period τ, along with periodical updates of the routing plan, with the same period τ.

After the warm-up time, both updates are still performed with the indicated period and based on information on the number of offered calls and effectively carried calls in the network for each flow $f_s, s \in \mathcal{S}$ gathered from real-time measurements, until the end of the simulation. Using this information, a calculation of $B(f_s)$ estimates, $s \in \mathcal{S}$ can be made, as well as a calculation of the values of all the upper and lower level o.f. related to blocking probabilities. As for the revenues, the knowledge of the effectively carried calls in the network allows for the calculation of the carried traffic estimates, hence the calculation of revenues follows straightforwardly.

For further information on the simulation platform used, see [13].

In table 3, the analytical values of each o.f. are displayed, together with the simulation results (average value ± half length of the 95% confidence interval) for these functions. For the static model simulation, different values for the warm-up time were tried, and the results displayed in table 3 were obtained with $t_{warm-up} = 8$h for a total simulation time of 48h. For the dynamic model simulation, different values for the times t_0 and t_1, for τ and for b were tried, and the results displayed in table 3 were obtained with $t_0 = t_1 = 4$h; $\tau = 10$m, 15m, 20m, 30m; $b = 0.3$; total simulation time of 48h. The revenue values have 2 decimal places and the blocking probability values have 3 significant figures.

Generally speaking, the analytical results obtained with HMOR-S2$_{\mathrm{PAS}}$ are not inside

Table 3: Average o.f. values, and 95% confidence intervals, for simulation of the static (considering a warm-up time of 8h) and the dynamic routing model (considering $t_0 = t_1 = 4$h, update period τ and $b = 0.3$), a total simulation time of 48h on the test network $\mathcal{M}$, for different values of α, when the HMOR-S2 was used to update the routing plan

α	O.f.	Initial solution	Analytical results	Static routing model results	Dynamic routing model results			
					$\tau = 10$m	$\tau = 15$m	$\tau = 20$m	$\tau = 30$m
0.0	W_Q	54803.69	64905.26*	64774.12±68.28	64776.24±76.03	64776.67±67.98	64774.68±68.46	64750.27±61.42
	$B_{Mm\|Q}$	0.413	0.0752	0.0773±0.00356	0.0774±0.00363	0.0770±0.00319	0.0771±0.00333	0.0793±0.00302
	$B_{m1\|Q}$	0.413	0.0752	0.0773±0.00356	0.0774±0.00363	0.0770±0.00319	0.0771±0.00333	0.0793±0.00302
	$B_{m2\|Q}$	0.314	0.0184	0.0236±0.000576	0.0235±0.000655	0.0236±0.000543	0.0237±0.000601	0.0238±0.000696
	$B_{m3\|Q}$	0.0198	0.00184	0.00200±0.0000499	0.00200±0.0000445	0.00200±0.0000525	0.00200±0.0000567	0.00204±0.0000592
	$B_{M1\|Q}$	0.912	0.708	0.706±0.00912	0.702±0.0145	0.704±0.0142	0.704±0.0115	0.682±0.0198
	$B_{M2\|Q}$	0.766	0.103	0.110±0.00600	0.108±0.0154	0.111±0.00501	0.111±0.00488	0.0916±0.00800
	$B_{M3\|Q}$	0.0585	0.0301	0.0303±0.000146	0.0300±0.000287	0.0302±0.0000739	0.0303±0.000384	0.0292±0.000978
	W_B	15106.57	17039.20	17017.10±39.32	17030.28±57.28	17017.43±42.53	17015.51±39.62	17059.73±42.35
0.5	W_Q	51785.21	60739.76⋄	60676.12±61.43	60659.89±53.91	60674.82±67.88	60675.17±66.08	60287.27±57.93
	$B_{Mm\|Q}$	0.413	0.0278	0.0306±0.00145	0.0317±0.00146	0.0310±0.00146	0.0308±0.00149	0.0533±0.00158
	$B_{m1\|Q}$	0.413	0.0278	0.0306±0.00145	0.0317±0.00146	0.0310±0.00146	0.0308±0.00149	0.0533±0.00158
	$B_{m2\|Q}$	0.296	0.00230	0.00463±0.000355	0.00511±0.000556	0.00458±0.000729	0.00460±0.000674	0.0163±0.000714
	$B_{m3\|Q}$	0.0174	0.000857	0.000922±0.0000167	0.000904±0.0000179	0.000910±0.0000118	0.000912±0.0000156	0.00105±0.0000369
	$B_{M1\|Q}$	0.882	0.629	0.626±0.0196	0.622±0.0243	0.634±0.0144	0.628±0.0207	0.481±0.0280
	$B_{M2\|Q}$	0.722	0.00959	0.0158±0.00216	0.0166±0.00142	0.0153±0.00127	0.0155±0.00130	0.0552±0.00348
	$B_{M3\|Q}$	0.0517	0.0244	0.0245±0.000261	0.0239±0.000315	0.0245±0.000259	0.0244±0.000236	0.0171±0.000325
	W_B	13787.49	16685.60	16696.08±40.87	16757.72±80.12	16710.52±74.37	16702.89±72.29	17562.04±41.26
1.0	W_Q	49010.41	56106.51•	56036.04±45.53	56044.88±66.47	56038.57±43.83	56036.66±49.07	55895.41±61.80
	$B_{Mm\|Q}$	0.405	0.0256	0.0274±0.00174	0.0271±0.00319	0.0272±0.00117	0.0273±0.00121	0.0375±0.00219
	$B_{m1\|Q}$	0.405	0.0256	0.0274±0.00174	0.0271±0.00319	0.0272±0.00117	0.0273±0.00121	0.0375±0.00219
	$B_{m2\|Q}$	0.275	0.00499	0.00805±0.000619	0.00772±0.00123	0.00796±0.000768	0.00803±0.000747	0.0131±0.00110
	$B_{m3\|Q}$	0.0150	0.000567	0.000643±0.0000157	0.000590±0.0000498	0.000635±0.0000189	0.000640±0.0000154	0.000686±0.0000270
	$B_{M1\|Q}$	0.841	0.556	0.552±0.0304	0.495±0.0826	0.555±0.0360	0.555±0.0262	0.354±0.0517
	$B_{M2\|Q}$	0.667	0.0186	0.0310±0.00318	0.0298±0.00504	0.0308±0.00347	0.0310±0.00446	0.0492±0.00497
	$B_{M3\|Q}$	0.0446	0.0200	0.0201±0.000295	0.0168±0.00444	0.0198±0.000717	0.0202±0.000248	0.0110±0.000972
	W_B	12445.64	16465.58	16436.45±17.45	16443.61±81.71	16434.68±23.36	16438.56±16.88	16690.18±50.64

*)99.62%; ⋄) 99.85%; •) 99.59% of the upper bounds $W_Q^{\max}$ for the optimal value of the QoS traffic revenue W_Q obtained in [21]

the 95% confidence interval of the static routing model simulation results, although they are of similar magnitudes. This is especially noticeable in situations of lower traffic loads (i.e. for higher values of α in our routing problem application example). In fact, only for $\alpha = 0.0$ did we get a result where a first level o.f. ($B_{Mm|Q}$) analytical value was in the corresponding confidence interval. The analytical results tend to be better than the corresponding static routing model simulation results, as could be expected. The differences between the analytical and the simulation results for the static routing model are mainly due to the imprecision effects intrinsic to the analytic/numerical solution, namely those associated with the simplifications of the traffic model, and the associated error propagation. In fact, the analytical model is a simplification which tends to underestimate the blocking probabilities in the network (and to overestimate the revenues), because the overflow traffic is treated as Poisson traffic, an error that is propagated throughout the complex and lengthy numerical calculations associated with the solution of the traffic model, involving the solution of large systems of implicit non-linear equations. Also in the stochastic model for the traffic in the links, a superposition of independent Poisson flows and an independent occupation of the links were assumed as further simplifications. A more precise and realistic representation of the traffic flows would allow for better estimates of the blocking probabilities. However, as mentioned earlier, the approximations in our model can be deemed appropriate and unavoidable in this context and result from a compromise between the precision of the representation of the traffic flows and the computational burden of the numerical resolutions throughout the execution of the heuristic algorithm.

The results of the static and dynamic routing models for $b = 0.3$ and $\tau = 10$m; 15m; 20m are quite similar. Considering all the simulation experiments performed, it was only for these parameter values that the dynamic routing model was capable of attaining the performance values corresponding to the analytic upper bounds for the static solution. Still, for $\alpha = 0.5$, the dynamic routing model results for the first level o.f. are slightly worse than those obtained with the static model. Remember that the simulation results for the dynamic routing model are average values of performance in a great number of routing update intervals, while the analytical results are obtained in ideal steady state traffic conditions. A final remark on the confidence intervals for each o.f.: their length is of the same order for both the static and the dynamic routing model.

In global terms and as expected, the results obtained with the dynamic routing model are better (or approximately the same in the worst case) than those obtained with the static routing model. This shows that the dynamic model is well calibrated for these networks, in terms of the choice of the initial routing solutions to be used by the heuristic and the choice of the routing updating period. In the dynamic routing model, the routing plan is adjusted throughout the simulation run, in accordance with the traffic random fluctuations around the average values corresponding to the nominal traffic matrix that was defined in steady state conditions.

The o.f. values are intrinsicly imprecise, due to the simplifications and approximations that were assumed in the stochastic model for the traffic in the links, and to the numerical errors, associated with the resolution of the system of equations of the traffic model which propagate throughout the resolution procedure. Still, the representation of the traffic flows as independent Poisson processes and the independence in the occupations of the links may be considered a good compromise between the exactness of the traffic model and the computational burden for solving the analytical model. Plus, the blocking probabilities in the system are calculated by 'robust' numerical algorithms (mentioned in the previous section) to try and reduce these imprecision effects.

Other imprecision effects are due to an instability phenomenon which may potentially arise in the path selection procedure, when all the network routes are liable to change. To avoid oscillations between certain solutions that can possibly lead to a poor global network performance, the core algorithm to seek new routing solutions (MMRA-S2) is applied only to specific traffic flows that are carefully chosen, so as to try and improve the o.f. values.

The inaccuracy in the o.f. values may be considered inevitable, due to the strong interdependence of the o.f. parameters and the (discrete) decision variables $\overline{R}$ (network route plans). The heuristic resolution has different mechanisms throughout the improvement cycles of the o.f. that try to deal with this complex interdependence, in order to minimise the impact of these inaccuracies in the quality of the obtained compromise solutions.

As mentioned earlier, the stochastic nature of the traffic offered to the network leads to some uncertainty in the results. In particular, for the state-dependent periodic-type dynamic routing model, the traffic flow means are periodically updated (with period τ) via a statistical estimate (first-order moving average) based on real-time measurements, dependent on a parameter b. The influence of these parameters on the final global routing solution was analysed.

In a first set of experiments, the update period τ was 15m and different values were tested for b. A sample of the results for $\alpha = 0.0$ is displayed in table 4. The parameter b has to be set in order to reflect a compromise between the stability of the estimate and the quick response to variations in the partial estimate of the average value of the traffic offered by a flow to the network in the previous interval, $\tilde{X}$. The best results for the first level o.f. were obtained with $b = 0.3$ (those are the values displayed in table 3). For $b = 0.4$ the results were only slightly worse. However, for smaller values and for higher values of b, the results for W_Q and $B_{Mm|Q}$ were worse than those displayed in table 3. An increase(decrease) in b means that the estimate of the average traffic offered to the network by a flow, $\tilde{x}$, incorporates more(less) information on the previous interval and less(more) on the estimates of traffic that have been obtained throughout the duration of the experiment. These results show that a balance between these two aspects is clearly desirable and no excessive weight should be attributed to either of them. However, as the best results were obtained with $b < 0.5$, it appears that the stability of the estimate is slightly more important than a rapid response to variations in the offered traffic. Notice that the possibility of network changes or sudden strong alterations in traffic patterns are not being considered in this study. This type of events would have required the traffic estimate to be able to respond better to very rapid variations in the offered traffic, so a value of $b > 0.5$ would have been expected to be more appropriate in that situation. Probably that would have been the case if the network traffic was not modelled as Poisson traffic, but rather as self-similar traffic, with its typical bursts of traffic (see for instance [18]), a situation out of the scope of the present study.

In a second set of experiments, the parameter b was kept at 0.3 and different values were tested for the update period τ. The best results for the first level o.f. were obtained with τ = 10m, 15m, 20m (these values are displayed in table 3). For a smaller value of τ (5m), the results for W_Q and $B_{Mm|Q}$ were slightly worse than these. However, for a higher value of τ (30m), the results were clearly worse (see table 3). Therefore, there is no need to update the traffic estimate too often (with update periods of the order of only a few minutes), but a very long update period is also undesirable. Notice that a value of $b = 0.3$ means that the update information focuses more on the medium/long

Table 4: Average o.f. values, and 95% confidence intervals, for simulation of the static (considering a warm-up time of 8h) and the dynamic routing model (considering $t_0 = t_1 = 4$h, update period $\tau = 15$m and different values of b), a total simulation time of 48h on the test network $\mathcal{M}$, for $\alpha = 0.0$, when the HMOR-S2 was used to update the routing plan

O.f.	Dynamic routing model results for $\alpha = 0.0$ and update period $\tau = 15$m			
	$b = 0.2$	$b = 0.3$	$b = 0.4$	$b = 0.5$
W_Q	64766.82±57.11	64776.67±67.98	64774.10±63.62	64769.62±64.25
$B_{Mm\|Q}$	0.0780±0.00353	0.0770±0.00319	0.0777±0.00304	0.0781±0.00278
$B_{m1\|Q}$	0.0780±0.00353	0.0770±0.00319	0.0777±0.00304	0.0781±0.00278
$B_{m2\|Q}$	0.0236±0.000432	0.0236±0.000543	0.0235±0.000618	0.0235±0.000807
$B_{m3\|Q}$	0.00200±0.0000401	0.00200±0.0000525	0.00199±0.0000427	0.00200±0.0000410
$B_{M1\|Q}$	0.694±0.0159	0.704±0.0142	0.698±0.00918	0.691±0.0166
$B_{M2\|Q}$	0.103±0.0137	0.111±0.00501	0.106±0.00756	0.0993±0.0110
$B_{M3\|Q}$	0.0299±0.000512	0.0302±0.0000739	0.0297±0.000143	0.0298±0.000415
W_B	17014.02±42.39	17017.43±42.53	17014.40±40.83	17022.10±30.93

term rather than on the short term, as the estimate of the average traffic offered to the network by a flow incorporates more information on the estimates of traffic that have been obtained throughout the duration of the experiment, rather than on the previous interval. In very long intervals slight changes in the offered traffic pattern are more likely to occur and these would tend to be disregarded, if the update periods were very long (30 min or more) because of the minor importance attached to the information on the specific previous interval, in the traffic estimate.

5 Conclusions

In this report, a stochastic two-level hierarchical multiobjective routing model for MPLS networks with two service classes and alternative routing was reviewed and analysed concerning sources of IIU. Key issues raised by its great complexity were discussed, as well as the major factors that constitute the sources of imprecision, inaccuracy and uncertainty of the model. The mechanisms used by the developed resolution heuristic approach to deal with these issues were described.

Analytical and stochastic simulation experiments (both static and dynamic) with the developed procedure were carried out, enabling the evaluation of inaccuracies intrinsic to the model and to the analytic/numerical resolution method. The possible effects of these inaccuracies on the results of the heuristic resolution procedure were discussed, as well as the forms of minimising their impacts on the heuristic effectiveness.

Furthermore, the experimental study, using discrete-event stochastic simulation, enabled the validation of the routing model results and the evaluation of effects in the model results of the uncertainty associated with the offered traffic estimates, in a dynamic version of the routing method.

References

[1] D. Bouyssou. Modelling inaccurate determination, uncertainty, imprecision using multiple criteria. In A. G. Lockett and G. Islei, editors, *Improving Decision Making in Organisations*, volume 335 of *LNEMS*, pages 78–87. Springer, 1989.

[2] J. Clímaco and J. Craveirinha. Multicriteria analysis in telecommunication network planning and design – Problems and issues. In J. Figueira, S. Greco, and M. Ehrgott, editors, *Multiple Criteria Decision Analysis – State of the Art Surveys*, volume 78 of *International Series in Operations Research & Management Science*, pages 899–951. Springer, 2005.

[3] J. C. N. Clímaco, J. M. F. Craveirinha, and M. M. B. Pascoal. An automated reference point-like approach for multicriteria shortest path problems. *Journal of Systems Science and Systems Engineering*, 15(3):314–329, Sep. 2006.

[4] J. C. N. Clímaco, J. M. F. Craveirinha, and M. M. B. Pascoal. Multicriteria routing models in telecommunication networks – Overview and a case study. In Y. Shi, D. L. Olson, and A. Stam, editors, *Advances in Multiple Criteria Decision Making and Human Systems Management: Knowledge and Wisdom*, pages 17–46. IOS Press, 2007.

[5] J. Craveirinha, R. Girão-Silva, and J. Clímaco. A meta-model for multiobjective routing in MPLS networks. *Central European Journal of Operations Research*, 16(1):79–105, Mar. 2008.

[6] J. Craveirinha, R. Girão-Silva, J. Clímaco, and L. Martins. A hierarchical multi-objective 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.

[7] J. Craveirinha, R. Girão-Silva, J. Clímaco, and L. Martins. A hierarchical multiobjective routing model for MPLS networks with two service classes. In A. Korytowski, K. Malanowski, W. Mitkowski, and M. Szymkat, editors, *Revised Selected Papers of the 23rd IFIP TC7 Conference on System Modeling and Optimization, Cracow, Poland, July 23-27, 2007*, volume 312 of *IFIP Advances in Information and Communication Technology*, pages 196–219. Springer, 2009.

[8] J. Craveirinha, L. Martins, and J. N. Clímaco. Dealing with complexity in a multi-objective 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.

[9] J. Craveirinha, L. Martins, T. Gomes, C. H. Antunes, and J. N. Clímaco. A new multiple objective dynamic routing method using implied costs. *Journal of Telecommunications and Information Technology*, 3:50–59, 2003.

[10] H. M. Elsayed, M. S. Mahmoud, A. Y. Bilal, and J. Bernussou. Adaptive alternate-routing in telephone networks: Optimal and equilibrium solutions. *Information and Decision Technologies*, 14:65–74, 1988.

[11] R. Girão-Silva, J. Craveirinha, and J. Clímaco. Hierarchical multiobjective routing in MPLS networks with two service classes - A meta-heuristic solution. *Journal of Telecommunications and Information Technology*, 3:20–37, 2009.

[12] R. Girão-Silva, J. Craveirinha, and J. Clímaco. Hierarchical multiobjective routing in MPLS networks with two service classes - Report on a Pareto archived strategy. Research Report 18/2009 (ISSN 1645-2631), INESC-Coimbra (`www.inescc.pt`), Dec. 2009.

[13] R. Girão-Silva, J. Craveirinha, and J. Clímaco. Hierarchical multiobjective routing in Multiprotocol Label Switching networks with two service classes – A heuristic solution. *International Transactions in Operational Research*, 16(3):275–305, May 2009.

[14] J. S. Kaufman. Blocking in a shared resource environment. *IEEE Transactions on Communications*, COM-29(10):1474–1481, Oct. 1981.

[15] F. P. Kelly. Blocking probabilities in large circuit-switched networks. *Advances in Applied Probability*, 18(2):473–505, Jun. 1986.

[16] F. P. Kelly. Routing in circuit-switched networks: Optimization, shadow prices and decentralization. *Advances in Applied Probability*, 20(1):112–144, Mar. 1988.

[17] J. Knowles, M. Oates, and D. Corne. Advanced multi-objective evolutionary algorithms applied to two problems in telecommunications. *BT Technology Journal*, 18(4):51–65, Oct. 2000.

[18] W. Leland, M. Taqqu, W. Willinger, and D. Wilson. On the self-similar nature of Ethernet traffic (Extended version). *IEEE/ACM Transactions on Networking*, 2(1):1–15, 1994.

[19] L. Martins, J. Craveirinha, J. N. Clímaco, and T. Gomes. Implementation and performance of a new multiple objective dynamic routing method for multiexchange networks. *Journal of Telecommunications and Information Technology*, 3:60–66, 2003.

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

[21] D. Mitra and K. G. Ramakrishnan. Techniques for traffic engineering of multiservice, multipriority networks. *Bell Labs Technical Journal*, 6(1):139–151, Jan. 2001.

[22] J. W. Roberts. Teletraffic models for the Telecom 1 integrated services network. In *Proceedings of 10th International Teletraffic Congress*, Montreal, Canada, 1983.

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

[24] A. P. Wierzbicki. Reference point methods in vector optimization and decision support. Interim Report IR-98-017, International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria, Apr. 1998.

A Specification of the Notation Used in the Model

Table 5: Notation used in the model

O.f. calculation	
$W_{Q(B)} = \sum_{s\in\mathcal{S}_{Q(B)}} A^c_s w_s$	total expected network revenue associated with QoS(BE) traffic flows
$B_{Mm\mid Q} = \max_{s\in\mathcal{S}_Q}\{B_{ms}\}$	maximal average blocking probability among all QoS service types
$B_{ms\mid Q} = \frac{1}{A^o_s}\sum_{f_s\in\mathcal{F}_s} A(f_s)B(f_s)$	mean blocking probabilities for flows of type $s \in \mathcal{S}_Q$
$B_{Ms\mid Q} = \max_{f_s\in\mathcal{F}_s}\{B(f_s)\}$	maximal blocking probability defined over all flows of type $s \in \mathcal{S}_Q$

Blocking probabilities calculation	
$B(f_s)$	node to node blocking probability for all flows $f_s \in \mathcal{F}_s$
$B_{ks} = \mathcal{B}_s\left(\overline{d_k}, \overline{\rho_k}, C_k\right)$	blocking probabilities for micro-flows of service type s in link l_k
$\mathcal{B}_s$	basic function (implicit in the teletraffic analytical model) to calculate B_{ks}

Decision variables	
$\overline{R} = \cup_{s=1}^{\lvert\mathcal{S}\rvert} R(s)$	network routing plans
$R(s) = \cup_{f_s\in\mathcal{F}_s} R(f_s), s \in \mathcal{S}_Q \cup \mathcal{S}_B$	set of all the feasible routes for the traffic flows of type s
$R(f_s) = (r^p(f_s)), p = 1, \cdots, M$	first, second, $\cdots$, M-th choice route for flow f_s

Path metrics and auxiliary parameters – MMRA-S2	
$m^1_{ks} = c^{Q(B)}_{ks}$	marginal implied costs
$m^2_{ks} = -\log(1 - B_{ks})$	marginal blocking probabilities
$\mathcal{D}(f_s)$	set of all feasible loopless paths for flow f_s

Preference thresholds used to define priority regions
$W_{req} = W_{av} + \Delta W = \frac{W_{av}+W_{Max}}{2}$
$W_{ac} = W_{av} - \Delta W = \frac{3W_{av}-W_{Max}}{2}$
$B^{log}_{req} = -\ln\left(1 - (B_{av} - \Delta B)\right) = -\ln\left(1 - \frac{B_{av}+B_{min}}{2}\right)$
$B^{log}_{ac} = -\ln\left(1 - (B_{av} + \Delta B)\right) = -\ln\left(1 - \frac{3B_{av}-B_{min}}{2}\right)$

Auxiliary parameters for the preference thresholds calculation
$W_{min} = \min_{X\in \text{ archive}} W_Q(X)$
$W_{Max} = \max_{X\in \text{ archive}} W_Q(X)$
$W_{av} = \frac{\sum_{X\in \text{ archive}} W_Q(X)}{\#\{X\}}$
$\Delta W = \frac{W_{Max}-W_{av}}{2}$
$B^{log}_{Mm\mid Q} = -\ln\left(1 - B_{Mm\mid Q}\right)$
$B_{min} = \min_{X\in \text{ archive}} B_{Mm\mid Q}(X)$
$B^{log}_{min} = -\ln\left(1 - B_{min}\right)$
$B_{Max} = \max_{X\in \text{ archive}} B_{Mm\mid Q}(X)$
$B^{log}_{Max} = -\ln\left(1 - B_{Max}\right)$
$B_{av} = \frac{\sum_{X\in \text{ archive}} B_{Mm\mid Q}(X)}{\#\{X\}}$
$\Delta B = \frac{B_{av}-B_{min}}{2}$
X is a solution in the archive

Table 5: Notation used in the model (continued)

Information on the priority regions	
$\mathcal{R}$	priority region in the o.f. space
ϱ	solution in $\mathcal{R}$
$\mathcal{C}_1(\varrho) = B^{log}_{Mm\vert Q}(\varrho)$	first metric for solution ϱ
$\mathcal{C}_2(\varrho) = W_Q(\varrho)$	second metric for solution ϱ
$\left(\mathcal{C}^*_{1\vert\mathcal{R}};\mathcal{C}^*_{2\vert\mathcal{R}}\right) = \left(m_{B^{log}\vert\mathcal{R}};M_{W\vert\mathcal{R}}\right)$	reference point in $\mathcal{R}$
$m_{i\vert\mathcal{R}}$	minimum of metric i for region $\mathcal{R}$
$M_{i\vert\mathcal{R}}$	maximum of metric i for region $\mathcal{R}$
$w_{i\vert\mathcal{R}} = \frac{1}{M_{i\vert\mathcal{R}}-m_{i\vert\mathcal{R}}}$	weight of metric i and region $\mathcal{R}$ in the weighted Chebyshev distance
$\left\{w_{i\vert\mathcal{R}}\left\vert\mathcal{C}_i(\varrho)-\mathcal{C}^*_{i\vert\mathcal{R}}\right\vert\right\}$	Chebyshev metric for metric i and region $\mathcal{R}$

Simulation parameters	
$T = [T_{ij}]$	base matrix with offered bandwidth values from node i to node j [Mbps]
α	compensation parameter
t_0	duration of the first stage of the initialisation phase, where only periodical updates of the estimate of the offered traffic are performed
t_1	duration of the second stage of the initialisation phase, where periodical updates of the estimate of the offered traffic and of the routing plan are performed
$t_{warm-up} = t_0 + t_1$	duration of the initialisation phase
τ	update period of the estimates of the offered traffic and of the network routing plans
$\tilde{x}_n(f_s) = (1-b)\tilde{x}_{n-1}(f_s) + b\tilde{X}_{n-1}(f_s)$	estimate of the average traffic offered to the network by the flow f_s in the time interval $[n\tau;(n+1)\tau[$
$\tilde{X}_{n-1}(f_s)$	estimator of the average value of the traffic offered by f_s to the network in the previous interval $[(n-1)\tau;n\tau[$
b	compromise value between the need to obtain a quick response of the estimator to rapid fluctuations in $\tilde{X}(f_s)$ and the stability of the long-run variations

Miscellany of auxiliary parameters	
f_s	flow of service type s
$\mathcal{S}_{Q(B)}$	set of QoS(BE) service types
A^o_s	total traffic offered by flows of type s
A^c_s	carried traffic for service type s
$A(f_s)$	mean traffic offered associated with $f_s \in \mathcal{F}_s$

Table 5: Notation used in the model (continued)

Miscellany of auxiliary parameters	
w_s	expected revenue per call of service type s
ρ_{ks}	reduced traffic loads offered by flows of type s to l_k
$\overline{\rho_k} = \left(\rho_{k1}, \cdots, \rho_{k\|\mathcal{S}\|}\right)$	vector of reduced traffic loads
d_{ks}	equivalent effective bandwidths for flows of type s in l_k
$\overline{d_k} = \left(d_{k1}, \cdots, d_{k\|\mathcal{S}\|}\right)$	vector of equivalent effective bandwidths
d'_s	required bandwidth for service s [kbps]
$d_s = \frac{d'_s}{u_0}$	required effective bandwidth for service s [channels]
$\mathcal{M}$	test network
$\|\mathcal{N}\|$	number of nodes in the network
$\|\mathcal{L}\|$	number of unidirectional links in the network
C'_k	link bandwidth [Mbps]
$C_k = \left\lceil \frac{C'_k}{u_0} \right\rceil$	link capacity [channels]
u_0	basic unit capacity
h_s	average duration of a type s call
D_s	maximum number of arcs for a type s call
$\xi(f_s)$	function for choosing candidate paths for flow f_s for possible routing improvement