Available online at www.sciencedirect.com
ScienceDirect
Optical Switching and Networking 5 (2008) 107–122
Optical Switching and Networking
www.elsevier.com/locate/osn

Grade-of-service differentiated static resource allocation schemes in WDM networks

James Yiming Zhang[a,1], Jing Wu[b,*], Gregor v. Bochmann[a,1], Michel Savoie[b,2]

[a] School of Information Technology and Engineering, University of Ottawa, Ottawa, Ontario, Canada, K1N 6N5
[b] Communications Research Centre Canada, Ottawa, Ontario, Canada, K2H 8S2

Received 17 August 2007; received in revised form 2 January 2008; accepted 24 January 2008
Available online 9 February 2008

Abstract

This paper presents a study on the Grade-of-Service (GoS) differentiation of static resource allocation in lightpath routed WDM networks, where lightpath requests between node pairs are given. Each request is associated with a service grade. The goal is to maintain certain service levels for the requests of all grades. The service levels are measured in terms of their acceptance ratios. We solve this network optimization problem by adopting a penalty-based framework, in which network design and operation goals can be evaluated based on cost/revenue. We propose a static GoS differentiation model as one minimizing the total rejection and cost penalty, in which the rejection penalty reflects the revenue of accepting a request, and the cost penalty reflects the resource consumption of providing a lightpath to a request. Then, a solution based on the Lagrangian relaxation and subgradient methods is used to solve the proposed optimization problem. Three different application scenarios are presented: static GoS differentiation of requests between the same node pair, static GoS differentiation of requests between different node pairs, and an integration of static GoS differentiation into the network profit objective. The fairness issues and the impact of relative penalty factors are discussed to provide guidelines for network planning.

Keywords: Resource allocation; Wavelength routed WDM networks; Routing and wavelength assignment; Service grade; Service differentiation; Fairness issues; Network profit optimization

1. Introduction

Grade-of-Service (GoS) is important in the design of Wavelength Division Multiplexing (WDM) networks, since optical networks serve an increasing number of services, each having different requirements. As Quad-play (data, voice, video and mobile communications) and high-performance computing traffic converge to IP and optical networking architectures, the resource allocation schemes in WDM networks must be able to provide GoS for the mixed traffic, i.e., the Routing and Wavelength Assignment (RWA) algorithms have to consider diversified requirements of the Quad-play traffic, such as the fairness of the request acceptance and resource allocations.

Classical RWA algorithms do not consider the distribution of accepted requests for different node pairs and for different service grades. There is no

* Corresponding author. Tel.: +1 613 9982474; fax: +1 613 9908382.
E-mail addresses: yizhang@site.uottawa.ca (J.Y. Zhang), jing.wu@crc.ca (J. Wu), bochmann@site.uottawa.ca (G.v. Bochmann), michel.savoie@crc.ca (M. Savoie).
[1] Tel.: +1 613 5625800x6205.
[2] Tel.: +1 613 9982489.

1573-4277/$ - see front matter Crown Copyright © 2008 Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.osn.2008.01.003

differentiated service grade in existing RWA algorithms, since they assume that all requests have the same priority (grade). Static RWA algorithms are used to allocate resources when lightpath requests are given. These algorithms typically aim at minimizing the overall resource requirement to accommodate all requests, or maximizing the overall utilization of a network that serves a single grade of requests. The latter is modeled as one minimizing the average number of fibre hops, minimizing the average node-to-node transmission delay, and minimizing network congestion. For a review, see [1,2]. However, with the network evolution from single to multiple services, it is important to provide a controlled GoS. Service grades can be defined according to network management policies [3,4]. In short, we use the terminology "static GoS differentiation" standing for GoS differentiated static resource allocation. In this paper, we focus on point-to-point logic connections (i.e., lightpaths) in wavelength routed WDM networks.

The challenge in providing a controlled GoS is that certain service levels should be maintained for requests of every grade. Meanwhile, pursuing maximal overall revenue of a converged transport network becomes even more challenging than that in single-grade traffic situation. The differentiation of service grades adds a dimension to the classical RWA problem, which requires new resource allocation algorithms. Although a high-grade request should have a better chance to be accepted (i.e., a high-grade request should experience a higher service level than a low-grade one), by no means a low-grade request can only be accepted after all high-grade requests are accepted. Requests of different grades generally share the same pool of resources. After allocating resources for all high-grade requests, resources may not be sufficient to maintain the service level of the low-grade. In contrast, rejecting a small set of high-grade requests could make a lot of critical resources available for low-grade requests, thus service levels are maintained for every grade. As such, service differentiation trade-offs must be studied. Specifically, the selection of rejected high-grade requests, and its impact on the network operation objectives should be investigated.

Static and dynamic GoS differentiations use distinct mechanisms. We study the static GoS differentiation, where all requests must be handled together as a whole. In contrast, in the dynamic GoS differentiation, requests are handled one by one. From the algorithm design point of view, the searching space of the global optimal solution of a static resource allocation is several orders greater than that in the dynamic case. None of the existing dynamic GoS differentiation mechanisms can be readily adapted for the static case. Three types of mechanisms are used to provide the dynamic GoS differentiation: resource preservation for future high-grade requests, in which each grade has a set of pre-determined wavelengths [5–7]; different routing for different grades, in which a high-grade request has more candidate routes, and thus has a better chance to be routed through than a low-grade one [7]; and controlled preemption of low-grade lightpaths [7,8].

We model the static GoS differentiation using a penalty/price-based optimization formulation. Assigning a high rejection penalty to a request makes the request less likely to be rejected than others. By assigning a proper relative rejection penalty for each request, a desired GoS differentiation can be achieved. The basis of our formulation is that all objectives such as GoS, fairness, and load-balancing can be integrated into one ultimate goal of maximizing the profit, and then be evaluated in monetary values. In our previous work [9], we optimized static resource allocations for a single grade of service by maximizing the revenue, which was also modeled as one minimizing the rejection penalty. However, we did not discriminate lightpaths using different network resources, and did not study the service differentiation. Here, we propose a new formulation considering both GoS differentiation and resource consumption.

Fairness deserves more study in the static GoS differentiation. The term fairness can be defined in different ways. Due to the wavelength continuity constraint, a request between two distant nodes suffers from a higher blocking ratio than a request between two nearby nodes [10]. A. Szymanski, et al., investigated the fairness in dynamic GoS differentiation [7]. Their computation results show that although the majority of node pairs in the high-grade service perceive a low blocking ratio, a few node pairs in the high-grade service perceive a much higher blocking ratio than the average. The GoS service contract of these unfairly treated node pairs may be violated. In our previous work, we studied the fairness of a static RWA algorithm for a single grade of requests [9]. So far, no study has been conducted on the fairness issue of the static GoS differentiation in the literature.

Specifically, the major contributions of this paper are as follows:

1. Proposal of a static GoS differentiation model as one minimizing the rejection and cost penalty: The rejection penalty reflects the revenue of accepting a request, while the cost penalty reflects the resource consumption of providing a lightpath to a request;

2. Proposal and evaluation of the static GoS differentiation of multiple requests between the same node pair: A step-shape rejection penalty function is provided, and the impact of incremental rejection penalties on the overall optimization objective and fairness between different node pairs are evaluated;
3. Evaluation of performance trade-offs of static GoS differentiation of requests between different node pairs: A group of node pairs are identified to be of a distinct grade;
4. Integration of the cost factor into the static GoS differentiation model and computation results exhibiting the influence of cost factors.

This paper is organized as follows: In Section 2, the network model and assumptions are summarized. In Section 3, we propose a formulation of static GoS differentiation, followed by a solution based on the Lagrangian relaxation and subgradient methods presented in Section 4. In Section 5, we present a study on static GoS differentiation of requests between the same node pair. In Section 6, static GoS differentiation is applied to requests between different node pairs. Then, we propose an integration of static GoS differentiation into the network profit objective in Section 7. Conclusions are given in Section 8. In the appendix, we provide a heuristic algorithm to derive a feasible solution to the original problem from the solution to the Lagrangian dual problem.

In this paper, the following notations and variables are used:

$c_{i\lambda}$	the cost of using a wavelength converter of index λ at node i;
C_{sdn}	the cost of the resources used by s_{sdn};
$d_{ij\lambda}$	the cost of using $w_{ij\lambda}$ on e_{ij};
D	the number of node pairs that have requests, but are not assigned any lightpath;
D_{sdn}	the cost of routing s_{sdn} in the dual problem;
e_{ij}	the fibre link between node pair (i, j), $e_{ij} \in \mathcal{E}$;
E	the number of fibre links in the network;
$\mathcal{E}$	the set of all fibre links in the network;
$F_{i\lambda}$	the number of wavelength converters of index λ at node i;
g	the subgradient function of q with respect to its given variables;
$I_i(\lambda)$	the set of wavelengths, to which at node i an input lightpath of wavelength λ can be converted;
J	the primal function of the original problem;
J^*	the optimal value of the primal function;
L	the number of requests;
$\mathcal{L}$	the set of all lightpath requests;
N	the number of nodes in the network;
P_{sdn}	the penalty for rejecting s_{sdn};
P^{DG}	the rejection penalty for a request of the distinct grade, when differentiating requests between different node pairs;
P^{RG}	the rejection penalty for a request of the regular grade;
q	the dual function $q(\xi, \pi)$, defined as the infimum of the Lagrangian function;
q^*	the optimal value of the dual function;
$q^{(h)}$	the value of q after the hth iteration when updating the Lagrange multipliers;
q^U	an estimate of the optimal solution of q when updating the Lagrange multipliers;
s_{sdn}	the nth request between node pair (s, d), $s_{\text{sdn}} \in \mathcal{L}$;
SP_{sdn}	the decomposed request-level subproblem corresponding to s_{sdn};
v	the degree of wavelength conversion, defined as the number of possible output wavelengths of a wavelength converter for a given input wavelength;
$\mathcal{V}$	the set of all nodes in the network;
$w_{ij\lambda}$	the wavelength channel λ on e_{ij}, $0 < \lambda \le W$;
W	the number of wavelengths in the network;
α_{sdn}	a binary integer variable representing the admission status of s_{sdn}. α_{sdn} is equal to 1, if s_{sdn} is admitted; otherwise, it is equal to 0;
A	the variable set $\{\alpha_{\text{sdn}}\}$, representing the admission status of all requests;
$\delta_{ij\lambda}^{\text{sdn}}$	a binary integer variable representing the use of $w_{ij\lambda}$ by s_{sdn}. $\delta_{ij\lambda}^{\text{sdn}}$ is equal to 1, if $w_{ij\lambda}$ is used by s_{sdn}; otherwise, it is equal to 0;
Δ_{sdn}	the variable set $\left\{\delta_{ij\lambda}^{\text{sdn}}\right\}_{\text{sdn}}$, representing the wavelength assignment at all links for s_{sdn};
Δ	the variable set $\{\Delta_{\text{sdn}}\}$, representing the wavelength assignment for all lightpaths;
$\phi_{i,ab}^{\text{sdn}}$	a binary integer variable representing the use of a wavelength converter by s_{sdn} at node i to convert an incoming lightpath of wavelength a to an outgoing lightpath of wavelength b. $\phi_{i,ab}^{\text{sdn}}$ is equal to 1, if s_{sdn} uses such a converter; otherwise, it is equal to 0;
Φ_{sdn}	the variable set $\left\{\phi_{i,ab}^{\text{sdn}}\right\}_{\text{sdn}}$, representing the converter assignment at all intermediate nodes of s_{sdn};
Φ	the variable set $\{\Phi_{\text{sdn}}\}$, representing the converter assignment for all lightpaths;
μ	the convergence control parameter used in updating the Lagrange multipliers;
π	the vectors of the Lagrange multipliers $\{\pi_{i\lambda}\}$;

$\pi_{i\lambda}$ the Lagrange multipliers with respect to the relaxation of the converter amount constraint;
$\theta^{(h)}$ the step size of the hth iteration when updating the Lagrange multipliers;
ξ the vectors of the Lagrange multipliers $\{\xi_{ij\lambda}\}$;
$\xi_{ij\lambda}$ the Lagrange multipliers with respect to the relaxation of the exclusive wavelength channel usage constraint;
Z a composite vector of the Lagrange multipliers, $Z = (\xi, \pi)$;
$Z^{(h)}$ the value of Z obtained at the hth iteration when updating the Lagrange multipliers.

2. Network model and assumptions

We consider a general WDM mesh network of N nodes interconnected by E fibre links. Each link has a pair of fibres, one fibre for each direction. Each fibre has W non-interfering Wavelength Channels (WCs). In our model, WCs are uni-directional, which means all the WCs in a fibre can only carry signals traveling towards the same direction. So if there is a fibre between node i and j, then $n_{ij} = W$. Two nodes may be logically connected through a lightpath, defined as a concatenated sequence of WCs using the same wavelength. We allow more than one lightpath being set up between a node pair. It is allowed to chain two lightpaths of different wavelengths (colors) together by using a wavelength converter installed at an intermediate node. For simplicity, we also call the resulting logical connection a lightpath.

We use a model of a wavelength converter with a limited conversion degree. Some of the most promising all-optical wavelength translation techniques, such as four-wave mixing in Semiconductor Optical Amplifiers (SOAs), have strong relations between the input and output wavelengths [11]. The converters based on these technologies have some degree of wavelength dependency too, and thus we use the limited converter structure [9,12] to model these relationships. In our model, an incoming lightpath of wavelength c can use a converter of a certain type (called the converter of index c) to be converted to one of the wavelengths in set $I_i(c) = \{c, c+1, \ldots, (c + (v-1))\} \bmod W$. Note that other converter architectures can be formulated in a similar way by defining different $I_i(c)$. A limited number (possibly none) of converters with the same index are installed at a node in a share-per-node structure. This means that if a given input lightpath of wavelength c needs conversion, such conversion is only possible when one converter in the converter bank of index c is available [13].

Our traffic model assumes that point-to-point lightpath requests between all node pairs are given, and are fixed over time. The lightpath requests are represented by a matrix, where the horizontal axis represents the source node, the vertical axis represents the destination node, and a number in the matrix represents the number of the requested lightpaths between a given node pair. Furthermore, if multiple lightpaths are set up between a node pair, they are not restricted to take the same route. Lightpaths are independently requested between given (s, d) and (d, s), and their acceptance and routing are independent as well.

3. Formulation of static grade-of-service differentiation

We formulate the static GoS differentiation as a minimizing-penalty problem, where the rejection of requests, and the use of resources are penalized. When a request is rejected, certain potential revenue is lost. Thus, the rejection penalty is the amount of its potential revenue. On the other hand, when a request is accepted, its resource consumption is added as a penalty in the overall objective function. The resource consumption penalty is the cost of resources that is used by the lightpath provisioned for the request. In this way, our objective essentially becomes the profit maximization. By adjusting the rejection penalty, we can incorporate the fairness consideration.

Our objective function is formulated as follows:

$$\min_{A,\Delta,\Phi} \{J\}, \quad \text{with}$$

$$J \equiv \sum_{s_{\mathrm{sdn}} \in \mathcal{L}} \left[(1-\alpha_{\mathrm{sdn}}) P_{\mathrm{sdn}} + \alpha_{\mathrm{sdn}} C_{\mathrm{sdn}} \right]. \tag{1}$$

For every request s_{sdn}, either the penalty of rejecting it (P_{sdn}), or the penalty of using resources (C_{sdn}) to set up a lightpath for it is added to the overall penalty function (J), depending on s_{sdn}'s admission status α_{sdn}. The design variables are the admission status of all requests (A), the wavelength assignment for all lightpaths (Δ), and the converter assignment for all lightpaths (Φ). As an example, we define C_{sdn} as the cost of using WCs (represented by the first summation) and converters (represented by the second summation):

$$C_{\mathrm{sdn}} = \sum_{e_{ij} \in \mathcal{E}} \sum_{0<\lambda\le W} d_{ij\lambda} \delta_{ij\lambda}^{\mathrm{sdn}} + \sum_{i \in \mathcal{V}} \sum_{0<\lambda\le W} \sum_{0<a\le W} c_{i\lambda} \phi_{i,\lambda a}^{\mathrm{sdn}}. \tag{2}$$

The optimization is subject to the following constraints:

(a) Lightpath flow continuity constraint,

$$\sum_{j\in\mathcal{V}}\sum_{0<c\le W}\delta_{ijc}^{\mathrm{sdn}}-\sum_{j\in\mathcal{V}}\sum_{0<c\le W}\delta_{jic}^{\mathrm{sdn}}=\begin{cases}\alpha_{\mathrm{sdn}} & \text{if } i=s\\ -\alpha_{\mathrm{sdn}} & \text{if } i=d\\ 0 & \text{otherwise}\end{cases}\quad \forall s_{\mathrm{sdn}}\in\mathcal{L}. \tag{3}$$

If a lightpath is admitted, the lightpath must be continuous along its path, and terminates at its two end nodes. If s_{sdn} is accepted (i.e., $\alpha_{\mathrm{sdn}}=1$), we measure the incoming and outgoing flows that it contributes:

- At its source node, it contributes one outgoing flow;
- At its destination node, it contributes one incoming flow;
- At each of its intermediate nodes, it contributes one incoming and one outgoing flow, thus the contributed incoming and outgoing flows are balanced.
- At a node that s_{sdn} does not flow through or terminate at, it does not contribute incoming or outgoing flows.

If s_{sdn} is rejected, neither does it contribute incoming or outgoing flows at any node.

(b) Exclusive WC usage constraint,

$$\sum_{s_{\mathrm{sdn}}\in\mathcal{L}}\delta_{ij\lambda}^{\mathrm{sdn}}\le 1\quad \forall e_{ij}\in\mathcal{E},\ 0<\lambda\le W. \tag{4}$$

A WC may be used by not more than one lightpath.

(c) Wavelength conversion constraints,

$$\phi_{j,ab}^{\mathrm{sdn}}=\begin{cases}1 & \text{if } \exists m,k\in\mathcal{V} \text{ and } b\ne a,\\ & \delta_{mja}^{\mathrm{sdn}}=\delta_{jkb}^{\mathrm{sdn}}=1\\ 0 & \text{otherwise.}\end{cases} \tag{5}$$

A wavelength converter at an intermediate node j is used only when s_{sdn} are assigned to different incoming and outgoing wavelengths at the node.

(d) Limited wavelength conversion degree constraints,

$$\phi_{j,ab}^{\mathrm{sdn}}=0\quad \text{if } b\notin I_j(a). \tag{6}$$

A given incoming wavelength can only be converted to a certain set of wavelengths.

(e) Converter amount constraint,

$$\sum_{s_{\mathrm{sdn}}\in\mathcal{L}}\sum_{0<a\le W}\phi_{i,\lambda a}^{\mathrm{sdn}}\le F_{i\lambda}\quad \forall i\in\mathcal{V},\ 0<\lambda\le W. \tag{7}$$

The total number of the incoming lightpaths of wavelength λ at node i that use conversion is limited by the number of installed converters.

4. A solution based on the Lagrangian relaxation and subgradient methods

The complexity of the formulated problem is very high. For a network with N nodes, E links, W wavelengths and L requests, the problem defined in the previous section contains $|A|+|\Delta|+|\Phi|=L+E\times W\times L+N\times W$ binary integer variables, where $|\bullet|$ denotes the number of elements in the set. For example, a small network with $E=21$, $N=14$, $W=20$ and $L=268$, has 113,108 variables in total. Finding the exact optimum for a problem of this size is hardly possible for the computing facilities available today. Therefore, finding a suboptimal solution within a reasonable computation time is a practical choice, while knowing the proximity of the suboptimal solution to the real optimum will be an additional advantage. In this paper, we develop a solution method that applies the Lagrangian Relaxation (LR) and subgradient methods. Our method can find a suboptimal solution for fairly large networks, and meanwhile, the proximity of the suboptimal solution to the real optimum can be evaluated by a bound. In this paper, we will apply our method to a network with 28 nodes, 61 links, 32 WCs, and 568 requests, which leads to over 1.1 million design variables in total.

4.1. Decomposing the problem based on Lagrangian relaxation

The LR method is used to derive the Dual Problem (DP) of the original problem that we formulated. We choose to relax the exclusive WC usage constraint (4) and converter amount constraint (7). Accordingly, additional elements are added into the primal function J by using the corresponding Lagrange multipliers $\xi_{ij\lambda}$ and $\pi_{i\lambda}$. This leads to the following Lagrangian dual problem:

$$\begin{aligned}\max_{\xi,\pi\ge 0}(q)\le \min_{A,\Delta,\Phi}\Bigg\{&\sum_{s_{\mathrm{sdn}}\in\mathcal{L}}\Bigg[(1-\alpha_{\mathrm{sdn}})P_{\mathrm{sdn}}\\ &+\alpha_{\mathrm{sdn}}\Bigg(\sum_{e_{ij}\in\mathcal{E}}\sum_{0<\lambda\le W}d_{ij\lambda}\delta_{ij\lambda}^{\mathrm{sdn}}\\ &+\sum_{i\in\mathcal{V}}\sum_{0<\lambda\le W}\sum_{0<a\le W}c_{i\lambda}\phi_{i,\lambda a}^{\mathrm{sdn}}\Bigg)\Bigg]\\ &+\sum_{e_{ij}\in\mathcal{E}}\sum_{0<\lambda\le W}\xi_{ij\lambda}\Bigg(\sum_{s_{\mathrm{sdn}}\in\mathcal{L}}\delta_{ij\lambda}^{\mathrm{sdn}}-1\Bigg)\\ &+\sum_{i\in\mathcal{V}}\sum_{0<\lambda\le W}\pi_{i\lambda}\Bigg(\sum_{s_{\mathrm{sdn}}\in\mathcal{L}}\sum_{0<a\le W}\phi_{i,\lambda a}^{\mathrm{sdn}}-F_{i\lambda}\Bigg)\Bigg\},\end{aligned} \tag{8}$$

subject to the constraints (3), (5) and (6).

After re-grouping the relevant terms, the dual function leads to the following problem:

$$\min_{A,\Delta,\Phi}\left\{\sum_{s_{\mathrm{sdn}}\in\mathcal{L}}\left[(1-\alpha_{\mathrm{sdn}})P_{\mathrm{sdn}} + \alpha_{\mathrm{sdn}}\left(\sum_{e_{ij}\in\mathcal{E}}\sum_{0<\lambda\le W} d_{ij\lambda}\delta_{ij\lambda}^{\mathrm{sdn}} + \sum_{i\in\mathcal{V}}\sum_{0<\lambda\le W}\sum_{0<a\le W} c_{i\lambda}\phi_{i,\lambda a}^{\mathrm{sdn}}\right) + \sum_{e_{ij}\in\mathcal{E}}\sum_{0<\lambda\le W}\xi_{ij\lambda}\delta_{ij\lambda}^{\mathrm{sdn}} + \sum_{i\in\mathcal{V}}\sum_{0<\lambda\le W}\sum_{0<a\le W}\pi_{i\lambda}\phi_{i,\lambda a}^{\mathrm{sdn}}\right] - \sum_{e_{ij}\in\mathcal{E}}\sum_{0<\lambda\le W}\xi_{ij\lambda} - \sum_{i\in\mathcal{V}}\sum_{0<\lambda\le W}\pi_{i\lambda}F_{i\lambda}\right\}. \tag{9}$$

Note that $\delta_{ij\lambda}^{\mathrm{sdn}}$ is equal to 0, if $\alpha_{\mathrm{sdn}} = 0$. Similarly, $\phi_{i,\lambda a}^{\mathrm{sdn}}$ is equal to 0, if $\alpha_{\mathrm{sdn}} = 0$. Therefore, we have the following two important relations, which lead to the decomposition of the dual problem.

$$\delta_{ij\lambda}^{\mathrm{sdn}} = \alpha_{\mathrm{sdn}}\delta_{ij\lambda}^{\mathrm{sdn}} \quad \forall e_{ij}\in\mathcal{E},\ 0<\lambda\le W \tag{10}$$

$$\phi_{i,\lambda a}^{\mathrm{sdn}} = \alpha_{\mathrm{sdn}}\phi_{i,\lambda a}^{\mathrm{sdn}} \quad \forall i\in\mathcal{V},\ 0<\lambda\le W. \tag{11}$$

By using (10) and (11), we re-write (9) as:

$$\min_{A,\Delta,\Phi}\left\{\sum_{s_{\mathrm{sdn}}\in\mathcal{L}}\left\{(1-\alpha_{\mathrm{sdn}})P_{\mathrm{sdn}} + \alpha_{\mathrm{sdn}}\left[\sum_{e_{ij}\in\mathcal{E}}\sum_{0<\lambda\le W}\left(d_{ij\lambda}+\xi_{ij\lambda}\right)\delta_{ij\lambda}^{\mathrm{sdn}} + \sum_{i\in\mathcal{V}}\sum_{0<\lambda\le W}\sum_{0<a\le W}(c_{i\lambda}+\pi_{i\lambda})\phi_{i,\lambda a}^{\mathrm{sdn}}\right]\right\} - \sum_{e_{ij}\in\mathcal{E}}\sum_{0<\lambda\le W}\xi_{ij\lambda} - \sum_{i\in\mathcal{V}}\sum_{0<\lambda\le W}\pi_{i\lambda}F_{i\lambda}\right\}. \tag{12}$$

Since the last two terms are independent of the decision variables, the problem can be further simplified as:

$$\min_{A,\Delta,\Phi}\sum_{s_{\mathrm{sdn}}\in\mathcal{L}}\left\{(1-\alpha_{\mathrm{sdn}})P_{\mathrm{sdn}} + \alpha_{\mathrm{sdn}}\left[\sum_{e_{ij}\in\mathcal{E}}\sum_{0<\lambda\le W}\left(d_{ij\lambda}+\xi_{ij\lambda}\right)\delta_{ij\lambda}^{\mathrm{sdn}} + \sum_{i\in\mathcal{V}}\sum_{0<\lambda\le W}\sum_{0<a\le W}(c_{i\lambda}+\pi_{i\lambda})\phi_{i,\lambda a}^{\mathrm{sdn}}\right]\right\}. \tag{13}$$

The relaxed problem (13) can be further decomposed into subproblems, each of which corresponds to one request. Corresponding to s_{sdn}, the request-level subproblem is defined as follows (denoted as $\mathrm{SP}_{\mathrm{sdn}}$):

$$\min_{\alpha_{\mathrm{sdn}},\Delta_{\mathrm{sdn}},\Phi_{\mathrm{sdn}}}\left\{(1-\alpha_{\mathrm{sdn}})P_{\mathrm{sdn}} + \alpha_{\mathrm{sdn}}\left[\sum_{e_{ij}\in\mathcal{E}}\sum_{0<\lambda\le W}\left(d_{ij\lambda}+\xi_{ij\lambda}\right)\delta_{ij\lambda}^{\mathrm{sdn}} + \sum_{i\in\mathcal{V}}\sum_{0<\lambda\le W}\sum_{0<a\le W}(c_{i\lambda}+\pi_{i\lambda})\phi_{i,\lambda a}^{\mathrm{sdn}}\right]\right\}, \tag{14}$$

subject to the constraints (3), (5) and (6).

4.2. Solving the derived subproblem

To solve the subproblem derived in Section 4.1, we can re-write $\mathrm{SP}_{\mathrm{sdn}}$ in (14) as:

$$\min_{\alpha_{\mathrm{sdn}}}\left[(1-\alpha_{\mathrm{sdn}})P_{\mathrm{sdn}} + \alpha_{\mathrm{sdn}}\min_{\Delta_{\mathrm{sdn}},\Phi_{\mathrm{sdn}}}(D_{\mathrm{sdn}})\right], \tag{15}$$

subject to the constraints (3), (5) and (6). where D_{sdn} is defined as:

$$D_{\mathrm{sdn}} = \sum_{e_{ij}\in\mathcal{E}}\sum_{0<\lambda\le W}\left(d_{ij\lambda}+\xi_{ij\lambda}\right)\delta_{ij\lambda}^{\mathrm{sdn}} + \sum_{i\in\mathcal{V}}\sum_{0<\lambda\le W}\sum_{0<a\le W}(c_{i\lambda}+\pi_{i\lambda})\phi_{i,\lambda a}^{\mathrm{sdn}}. \tag{16}$$

The solution to the subproblem (15) consists of three steps:

(1) Construct a wavelength graph of the network. The wavelength graph consists of $N \times W$ vertices in a matrix-like structure, where each column corresponds to a node of the network, and each row corresponds to a wavelength. In the λth ($0 < \lambda \le W$) row, if $w_{ij\lambda}$ ($\forall e_{ij} \in \mathcal{E}$) exists, then draw a directed edge from the column i to j, and assign weight $\left(d_{ij\lambda}+\xi_{ij\lambda}\right)$ to the edge to represent the WC. In the column i ($\forall i \in \mathcal{V}$), if wavelength conversion is allowed from wavelength λ to a, (i.e., $a \in I_i(\lambda)$) at node i, then draw a directed edge

from the row λ to a, and assign weight $(c_{i\lambda} + \pi_{i\lambda})$ to the edge to represent the wavelength converter;
(2) Find a shortest path between node pair (s, d), and calculate the optimal value of $\min_{\Delta_{\text{sdn}}, \Phi_{\text{sdn}}} (D_{\text{sdn}})$;
(3) Determine the acceptance or rejection of the request s_{sdn}. If $P_{\text{sdn}} > \min_{\Delta_{\text{sdn}}, \Phi_{\text{sdn}}} (D_{\text{sdn}})$, then accept s_{sdn} (i.e., set $\alpha_{\text{sdn}} = 1$), and assign s_{sdn} with the RWA scheme that corresponds to the shortest path in the wavelength graph. If $P_{\text{sdn}} < \min_{\Delta_{\text{sdn}}, \Phi_{\text{sdn}}} (D_{\text{sdn}})$, then reject s_{sdn} (i.e., set $\alpha_{\text{sdn}} = 0$). A tie of P_{sdn} and $\min_{\Delta_{\text{sdn}}, \Phi_{\text{sdn}}} (D_{\text{sdn}})$ is broken arbitrarily.

To reduce the computation time, the shortest paths from a source node to all destination nodes can be computed all together at one time. The overall complexity of this algorithm is $O((N + W) NW)$. Compared to regular Dijkstra algorithm's complexity $O\left((NW)^2\right)$, lower complexity is achieved by using a special shortest path algorithm [14], which takes advantage of the special structure of the wavelength graph.

4.3. Updating the Lagrange multipliers

The subgradient method [15] is used to solve DP. The multiplier vector $Z = (\xi, \pi)$ is updated by the following formula:

$$Z^{(h+1)} = Z^{(h)} + \theta^{(h)} g\left(Z^{(h)}\right). \tag{17}$$

The $g(Z)$ is a composite vector of $g\left(\xi_{ij\lambda}\right)$ and $g(\pi_{i\lambda})$, which are computed as:

$$g(\xi_{ij\lambda}) = \sum_{s_{\text{sdn}} \in \mathcal{L}} \delta_{ji\lambda}^{\text{sdn}} - 1 \quad \forall e_{ij} \in \mathcal{E},\ 0 < \lambda \le W \tag{18}$$

$$g(\pi_{i\lambda}) = \sum_{s_{\text{sdn}} \in \mathcal{L}} \sum_{0<a\le W} \phi_{i,\lambda a}^{\text{sdn}} - F_{i\lambda}$$
$$\forall i \in \mathcal{V},\ 0 < \lambda \le W. \tag{19}$$

The step size $\theta^{(h)}$ is determined by,

$$\theta^{(h)} = \mu \times \frac{q^U - q^{(h)}}{g^T\left(Z^{(h)}\right) g\left(Z^{(h)}\right)}. \tag{20}$$

In general, we let q^U take on the best value of J obtained from the feasible resource allocation scheme. Based on the established knowledge from the operations research [16], we adjust the parameters adaptively to speed up the convergence as the iterations evolve. The values of μ and q^U are updated accordingly. Specifically, if the value of $q^{(h)}$ remains roughly the same for a certain number (denoted by x) of iterations, the value of μ is decreased by a factor $p < 1$; and if the value of $q^{(h)}$ keeps increasing for a certain number (denoted by y) iterations, the value of μ is increased by a factor $\frac{1}{p}$. From our experiments, fast convergence is obtained when $p = 0.95$, $x = 3$, and $y = 5$. The value of q^U is also updated if a lower J value is obtained.

4.4. Constructing a feasible resource allocation

We use a heuristic algorithm to derive a feasible RWA scheme to the original problem from the solution to DP. Generally, a solution to DP might be an infeasible resource allocation, because some constraints are relaxed when creating DP. The relaxed constraints (4) and (7) might be violated, although other constraints are respected, which is guaranteed by the formulation and the solution of SP_{sdn} in (15). In our heuristic algorithm, the requests are sorted based on heuristic rules. Then, a feasible RWA scheme is searched for each request, in which the search is guided by the solution to DP. Details of the heuristic algorithm are given in the appendix. In the worst case, the computational complexity of the heuristic algorithm is $O\left(L\,(NW)^2\right)$.

4.5. Evaluating the constructed resource allocation

We use the *duality gap* to evaluate the resource allocation that is constructed by the heuristic algorithm. This measure has been well studied, and is supported by various background theories about the LR approach. The upper and lower bounds of J^* (the optimal value of the primal function) can be estimated: its upper bound is the value of J corresponding to a feasible resource allocation; and its lower bound is q^* (the optimal value of the dual function). The difference $(J^* - q^*)$ is known as the *duality gap*. Moreover, an upper bound of the duality gap is $(J - q)$. We use the duality gap's upper bound as a measure of the suboptimality of a feasible resource allocation. Thus even without obtaining the exact optimum, we know that the distance of a suboptimal solution from the optimum is within a certain range.

5. Static GoS differentiation of requests between the same node pair

To provide requests with GoS differentiation, we assign an increasing rejection penalty for requests from low to high grades. Our algorithm ensures that between the same node pair, a low-grade request can only be accepted, when all the requests of higher grades are accepted. In this section, for convenience we assume that between node pair (s, d), the highest-grade request

Table 1
An increasing rejection penalty for requests from low to high grades leads to a tendency of a fair resource allocation

Node pair (1, 2)		Node pair (3, 4)		Total penalty for rejecting requests in the network	D
Number of rejected requests between node pair (1, 2)	Total penalty for rejecting requests between node pair (1, 2)	Number of rejected requests between node pair (3, 4)	Total penalty for rejecting requests between node pair (3, 4)		
0	0	4	340	340	1
1	40	3	210	250	0
2	110	2	110	**220**	**0**
3	210	1	40	250	0
4	340	0	0	340	1

is s_{sd1}, followed by s_{sd2}, and so on. We also assume that the requests between a node pair fill in high grades first with a grade having at most one request. For example, if a node pair has two requests, then one of the requests belongs to the highest grade, and the other belongs to the second highest grade. All node pairs are assumed using the same grade ladder. This means that if the highest-grade request of each node pair belongs to the same grade, so does the second highest-grade request of each node pair, and so on.

The key issue is how the increment of rejection penalty impacts on the overall optimization objective, and the fairness between node pairs. In this paper, fairness reflects how the acceptance of requests spreads over node pairs. In general, a high degree of fairness among node pairs is associated with accepting requests from high to low grades across all node pairs, and low variability of the number of accepted requests of the same grade among various node pairs. The definition of fairness depends on network operations. We measure the fairness as the number of node pairs that have requests, but are not assigned any lightpath (denoted as D). The smaller D is, the better fairness the resource allocation scheme demonstrates. Ideally, D is zero.

An increasing rejection penalty for requests from low to high grades leads to a tendency of a fair resource allocation. Please note that our fairness consideration is defined for the sets of different priorities, instead of the individual requests. This is illustrated in a simple example shown in Fig. 1 and Table 1. In this example, when the requests of the lowest two grades between both node pairs are rejected, the optimal solution is obtained. The following assumptions are made in this example: (1) Every link has 4 WCs; (2) There are 4 requests between node pair (1, 2); and 4 requests between node pair (3, 4); (3) Within the same node pair, the penalties for rejecting the 1st, 2nd, 3rd and 4th requests are set to 40, 70, 100 and 130, respectively.

A high-grade request is not guaranteed to be accepted before the acceptance of a low-grade request

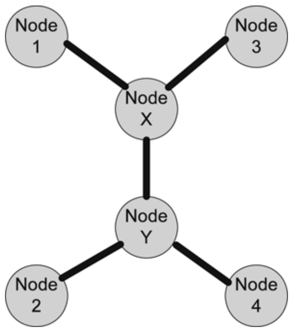

Fig. 1. An example for a 6-node network.

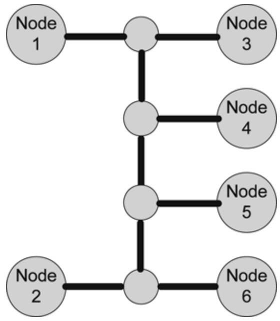

Fig. 2. An example for a 10-node network.

between a different node pair, although between the same node pair, the GoS differentiation is strictly respected. This is because routing lightpaths in a mesh topology creates a complicated competition for resources, and the increment of rejection penalty influences the fairness between node pairs. This is illustrated in another simple example shown in Fig. 2 and Tables 2 and 3. In this example, when the increment is set to 30 in scheme 1 (within the same node pair, the penalties for rejecting the 1st, 2nd, 3rd and 4th requests are set to 40, 70, 100 and 130, respectively), the optimal solution is one lightpath being allocated to node pair (1, 2). When the increment is reduced to 20 in scheme 2 (within the same node pair, the penalties for rejecting the 1st, 2nd, 3rd and 4th requests are set to 70,

Table 2
Penalty assignment scheme 1

Node pair (1, 2)		Node pair (3, 4)		Node pair (5, 6)		Total penalty for rejecting requests in the network	D
Number of rejected requests between node pair (1, 2)	Total penalty for rejecting requests between node pair (1, 2)	Number of rejected requests between node pair (3, 4)	Total penalty for rejecting requests between node pair (3, 4)	Number of rejected requests between node pair (5, 6)	Total penalty for rejecting requests between node pair (5, 6)		
0	0	4	340	4	340	680	1
1	40	3	210	3	210	460	0
2	110	2	110	2	110	330	0
3	210	1	40	1	40	**290**	**0**
4	340	0	0	0	0	340	2

Table 3
Penalty assignment scheme 2

Node pair (1, 2)		Node pair (3, 4)		Node pair (5, 6)		Total penalty for rejecting requests in the network	D
Number of rejected requests between node pair (1, 2)	Total penalty for rejecting requests between node pair (1, 2)	Number of rejected requests between node pair (3, 4)	Total penalty for rejecting requests between node pair (3, 4)	Number of rejected requests between node pair (5, 6)	Total penalty for rejecting requests between node pair (5, 6)		
0	0	4	400	4	400	800	1
1	70	3	270	3	270	610	0
2	160	2	160	2	160	480	0
3	270	1	70	1	70	410	0
4	400	0	0	0	0	**400**	**2**

90, 110 and 130, respectively), the optimal solution is all lightpaths being allocated to node pairs (3, 4) and (5, 6), none between node pair (1, 2). The following assumptions are made in this example: (1) Every link has 4 WCs; (2) There are 4 requests between node pair (1, 2); 4 requests between node pair (3, 4); and 4 requests between node pair (5, 6).

The increment of rejection penalty controls the fairness among node pairs, leading to a trade-off between the network revenue maximization and fairness. In the following example, two grades are used. For illustration, we assume each node pair has none, one or two requests. We use a network topology of 28 nodes (shown in Fig. 3). The traffic model is a randomly generated traffic matrix (shown in Table 4).

Our computations show that in a mesh network, as the grade differentiation becomes small, reduced fairness is observed. The number of disconnected node pairs (D) is shown with respect to the rejection penalty of a low-grade request in Figs. 4–6. The rejection penalty of a high-grade request is fixed at $P_{sd1} = 1000$. Our computations are conducted for different network traffic loads by setting a different number of WCs in the network. The number of WCs on each link is set to 14 (Fig. 4), 16 (Fig. 5) and 18 (Fig. 6), respectively. When the grade differentiation becomes small by increasing P_{sd2}, the rejection tends to be less fair, indicated by an increasing D. This is because the bigger the grade differentiation is, more resources are allocated to high-grade requests. As a result, each node pair tends to get its high-grade request accepted before resources are allocated to other requests. The differentiation between grades influences more on the fairness when the rejection ratio is high (at a smaller W) than when the rejection ratio is low. This is evidenced by a sharper increase of D in a heavy traffic load (Fig. 4) than in a medium or light traffic load (Figs. 5 and 6). However, the fairness improvement is achieved at the cost of rejecting more requests. This is because certain node pairs consume more resources than the average, and accepting their high-grade requests means sacrificing more low-grade requests between other node pairs, even though the latter node pairs consume fewer resources than the average.

6. Static GoS differentiation of requests between different node pairs

GoS differentiation can be applied to a certain group of node pairs. A typical example would be in the

Table 4
A randomly generated traffic matrix for the 28-node network

0	0	2	0	0	0	2	0	2	0	0	2	0	2	2	0	2	0	0	0	2	0	0	1	0	2	2	2
0	0	0	2	2	2	0	0	0	2	0	0	0	2	0	2	0	0	2	2	0	2	2	0	2	2	0	2
2	2	0	2	0	0	2	2	0	2	0	2	2	0	0	2	2	2	0	2	2	2	0	0	2	0	1	1
2	0	0	0	0	0	2	2	2	2	0	2	0	2	2	0	2	2	0	2	2	0	1	0	1	0	0	0
0	2	0	2	0	0	0	2	0	2	0	0	0	2	0	2	0	0	0	2	0	2	0	1	0	0	0	2
0	2	0	2	2	0	0	2	2	0	0	0	0	2	2	2	0	0	2	0	0	2	1	0	0	0	2	2
1	1	2	0	2	2	0	0	2	0	1	0	2	2	2	0	0	1	2	0	0	0	2	0	0	1	0	1
2	0	1	2	0	0	0	0	0	2	1	0	2	0	0	2	0	2	0	2	0	2	2	1	2	1	1	1
2	0	0	2	2	2	0	0	0	1	0	0	0	0	0	2	0	2	2	1	0	2	0	1	0	0	2	2
0	0	0	0	1	0	1	0	0	0	0	0	0	2	0	0	1	0	1	0	0	0	0	1	0	0	2	2
0	0	0	1	0	2	0	0	0	2	0	2	0	2	0	1	0	0	0	2	2	1	0	1	0	2	0	0
1	2	0	0	2	1	2	1	1	1	1	0	0	2	1	0	2	1	2	1	0	0	0	2	0	1	0	2
1	0	0	1	0	0	1	0	2	0	1	0	0	2	2	1	1	1	0	0	0	1	2	2	0	0	2	0
0	0	0	1	0	0	0	2	0	0	1	0	2	0	0	1	0	0	0	0	0	1	2	2	2	0	0	0
2	0	0	0	0	0	1	2	1	1	0	1	0	2	0	0	1	2	0	1	1	0	2	0	2	0	1	0
0	1	0	2	1	0	0	2	2	0	0	0	0	1	2	0	0	0	1	0	0	2	1	1	1	1	0	0
0	2	0	1	0	0	0	1	0	2	0	0	0	2	0	1	0	0	0	2	0	1	0	0	1	0	0	0
1	1	2	0	2	2	0	0	2	0	1	0	2	2	2	0	0	0	2	0	0	0	0	0	0	2	0	2
2	0	1	2	0	0	0	0	0	2	1	0	2	0	0	2	0	2	0	2	0	2	0	0	0	2	1	0
0	0	0	1	0	2	0	0	0	2	0	2	0	2	0	1	0	0	0	0	2	1	0	0	2	0	2	0
0	0	0	0	1	0	1	0	0	0	0	0	0	2	0	0	1	0	1	0	0	0	0	2	0	0	2	1
1	0	0	1	0	0	1	0	2	0	1	0	0	2	2	1	1	1	0	0	0	0	2	0	2	0	0	2
1	0	2	0	0	2	0	2	2	0	2	0	1	0	1	0	2	0	0	1	0	0	0	2	0	2	0	1
0	0	1	0	0	1	0	2	0	1	0	0	2	2	0	2	2	0	0	0	2	0	0	0	2	0	1	0
0	1	0	2	0	0	1	2	0	0	0	1	0	0	1	0	2	0	0	0	1	0	0	0	0	0	1	2
2	0	0	0	1	0	0	0	2	1	1	1	1	0	2	2	1	0	1	0	2	0	0	1	2	0	0	0
0	1	2	0	0	0	2	0	0	2	0	1	0	2	0	0	1	0	0	0	1	0	1	0	1	2	0	2
2	2	0	0	0	2	1	2	0	0	0	1	0	2	0	1	0	2	0	2	0	0	1	2	0	2	0	0

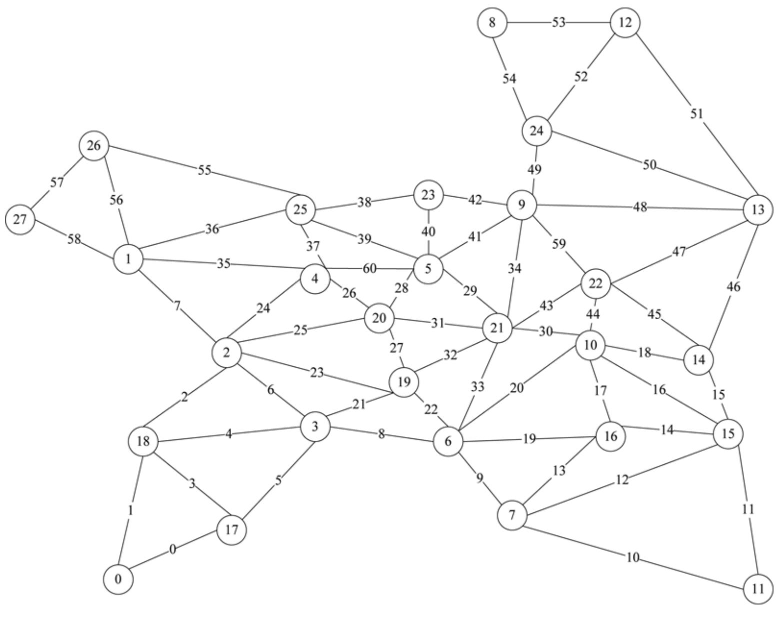

Fig. 3. Pan-European network with 28 nodes and 61 links.

resource allocation for a WDM Virtual Private Network (VPN), the requests between certain node pairs require a distinct GoS. We assume two grades are used to differentiate requests, a Distinct Grade (DG) and a Regular Grade (RG). Depending on network operation policy, a DG request may have a higher or lower grade

Table 5
A distinct grade (DG) mask for the 28-node network

0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	1	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	1	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
0	0	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
0	0	1	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	1	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0

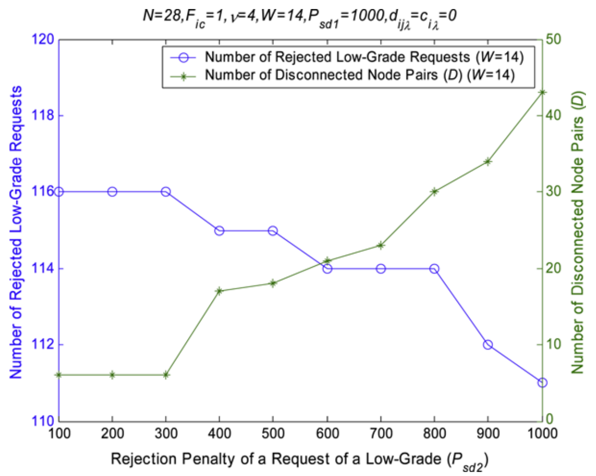

Fig. 4. The number of disconnected node pairs and the number of rejected low-grade requests with respect to the rejection penalty of a request of a low-grade (heavy traffic load case).

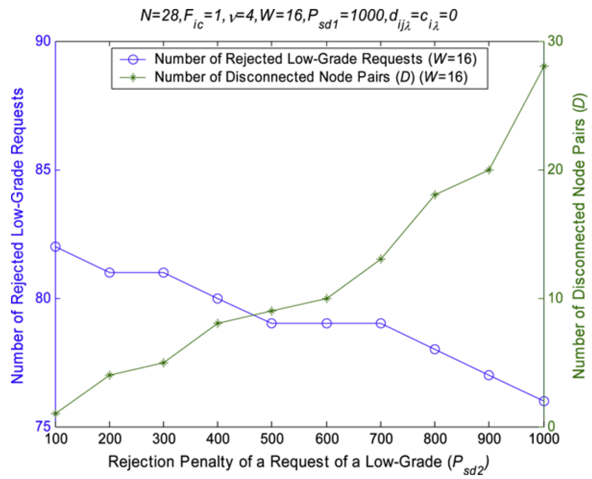

Fig. 5. The number of disconnected node pairs and the number of rejected low-grade requests with respect to the rejection penalty of a request of a low grade (medium traffic load case).

than an RG request. We use a DG mask to identify the differentiated node pairs. If node pair (s, d) belongs to DG, the cross of the sth row and the dth column is set to 1 in the mask. We use a randomly generated DG mask shown in Table 5. For illustration, we assign the rejection penalty for an RG request to 1000, and vary the rejection penalty for a DG request. We use the same network topology (Fig. 3) and traffic matrix (Table 4) given in the previous section. Applying the DG mask to the traffic matrix, we have 37 DG requests and 531 RG requests in total.

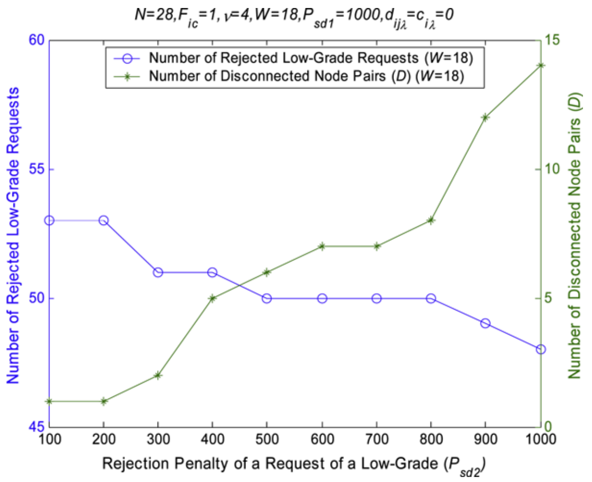

Fig. 6. The number of disconnected node pairs and the number of rejected low-grade requests with respect to the rejection penalty of a request of a low grade (light traffic load case).

There is a trade-off between accepting DG and RG requests (shown in Fig. 7). As P^{DG} increases, more DG requests are accepted, and meanwhile fewer RG requests are accepted. Initially when P^{DG} is less than 50, accepting DG requests does not generate profit, because the cost of a WC is set to 50. Thus all DG requests are rejected. When P^{DG} falls in the range between 50 and 200, the number of accepted DG requests increases drastically. As the rejection penalty of a DG request increases, more DG requests become profitable. At the same time, the number of accepted RG requests slowly decreases. This indicates that at this stage, the increased acceptance of DG requests is a result of their increased profit. Yet, their resource competition with RG requests does not have much influence. However, in the next stage where P^{DG} falls in the range between 200 and 1000, the number of accepted RG requests decreases drastically, while at the same time, the number of accepted DG requests also increases at a fast pace. This indicates that the increased acceptance of DG requests is a result of successful resource competition with RG requests. It should be noted that when the rejection penalty of a DG request happens to be identical to that of an RG request, the acceptance ratio of DG and RG requests is similar. In our example, their acceptance ratios are $32/37 = 0.86$ and $454/531 = 0.85$.

7. Integration of static GoS differentiation into the network profit objective

Fundamentally, the goal of static GoS differentiation is to maximize network profit. The rejection penalty

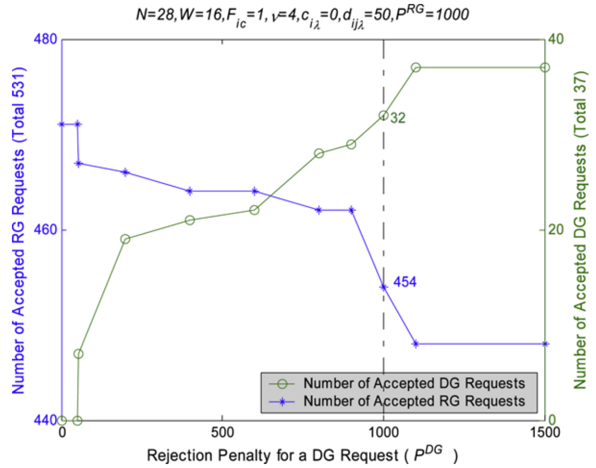

Fig. 7. Trade-off between accepting DG and RG requests.

of a request is in fact the price (i.e., revenue) of providing a lightpath to the request. In the previous two sections, we discussed that the relative price of different grades influences their acceptance. However, the cost of providing lightpaths also plays an important role in deciding the acceptance or rejection of requests. In this section, we consider the cost of the resources that lightpaths use. The challenge is that the required resources in providing service to a request highly depend on the overall resource usage and availability, which in turn depends on the acceptance of other requests. Thus the cost of a lightpath between a given node pair varies depending on the routing of the lightpath. Therefore, the decision of accepting or rejecting requests and associated resource allocations must be considered together, not separately.

In the first computation, we demonstrate the impact of $d_{ij\lambda}$ on the hop-count of lightpaths. In this computation, we do not differentiate requests and fix the price of lightpaths by assigning all P_{sdn}'s to the same value. As $d_{ij\lambda}$ decreases (shown in Fig. 8), lightpaths are able to take more hops (a hop is defined as going through one fibre link), and thus with the same amount of resources, more lightpaths can be accommodated. Since the revenue of a lightpath (P_{sdn}) is set to 1000, when $d_{ij\lambda} = 510$, no lightpath can be over two hops, because the revenue cannot cover the cost of two hops. The number of one or two-hop lightpaths hardly changes as $d_{ij\lambda}$ changes, because rerouting options are very limited for these lightpaths. However, for a lightpath with three or more hops, there is a variation as $d_{ij\lambda}$ changes, indicating trade-offs exist. For example, when $d_{ij\lambda}$ varies from 250 to 10, with little sacrifice of less lightpaths of two or three hops, our algorithm is able to accommodate many more lightpaths

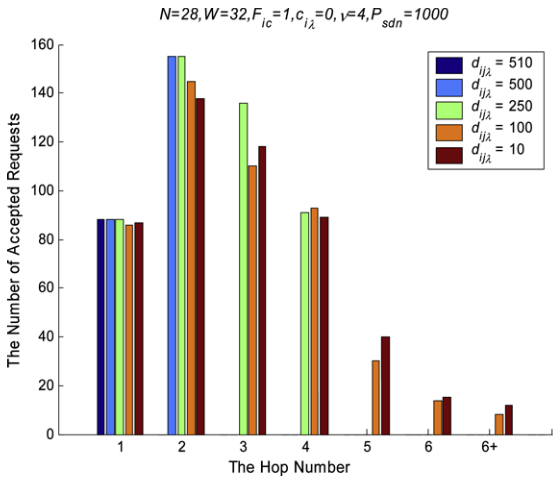

Fig. 8. Distribution of different hop-count lightpaths.

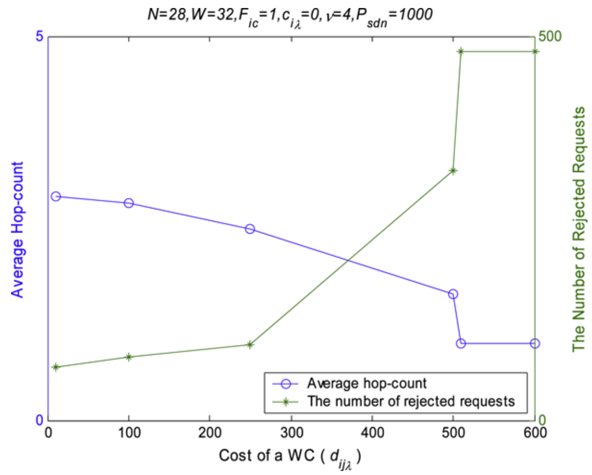

Fig. 9. Trade-off between average hop-count and the number of rejected requests.

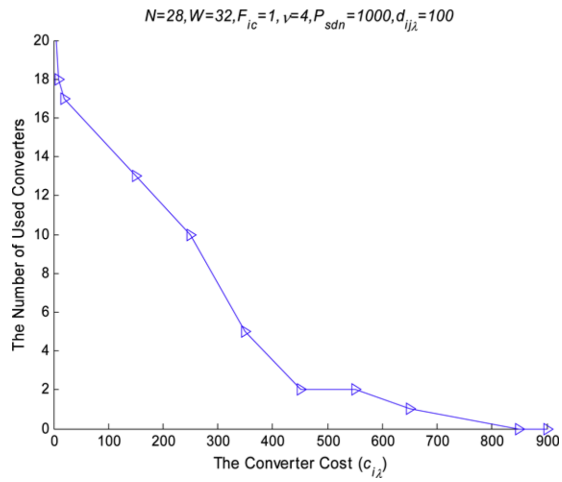

Fig. 10. The use of converters with respect to their costs.

of a larger hop-count. Through computations of other networks (results are not shown), we observe that the distribution of different hop-count lightpaths strongly depends on the nodal degree and the traffic load of the network. For a highly connected network (i.e., high nodal degree) under a heavy traffic load, when $d_{ij\lambda}$ is small to medium, more requests are accepted by using lightpaths with large (more than four) hops. The reason is twofold: on the one hand a highly connected network provides more options in routing a lightpath, and on the other hand, a heavy traffic load demands more resources, creating severe resource competition. However, regardless of network topology (we computed mesh networks with 14, 22 and 28 nodes) and traffic model, most of the lightpaths are routed within four hops. In Fig. 9, the trade-offs between the average hop number and the number of rejected requests are presented. As $d_{ij\lambda}$ increases, the average hop number drops, while the number of rejected requests increases. There is a sharp change when $d_{ij\lambda}$ reaches 250, because all the lightpaths having more than three hops cannot afford the resource cost any more. There is a steep descend when $d_{ij\lambda}$ reaches 500, after which no lightpath can afford more than a single hop.

We investigate the impact of the cost of converters on their usage. As their cost increases, their usage decreases, because the lightpaths that require wavelength conversion are more penalized. We quantitatively evaluate such a trend (shown in Fig. 10). When $c_{i\lambda}$ increases from 0 to 25, the use of converters drops most drastically, and after that, it decreases almost linearly with $c_{i\lambda}$. When it reaches 850, no converter is used. The reason is that when a lightpath uses a converter, the lightpath has at least two hops. Thus the cost of the lightpath is at least $c_{i\lambda} + 2d_{ij\lambda}$. In the example, after $c_{i\lambda}$ reaches 850, the cost of the lightpath is at least 1050. Since the revenue of the request is set to 1000, no profit can be generated from providing a lightpath using a converter to a request.

In the next example, we demonstrate the impact of resource cost on the acceptance of requests in the presence of GoS differentiation. We differentiate the requests between different node pairs as two grades: DG and RG (as in the previous section). As $d_{ij\lambda}$ increases, the rejection of DG and RG requests exhibits different behaviors (Fig. 11). Since a DG request has less revenue (in the computation, P^{DG} is set to 600, compared to P^{RG} being set at 1000), the rejection ratio of DG requests is approximately twice as much as that of RG requests, before all DG requests are rejected when resources are too expensive ($d_{ij\lambda}$ is too high). The

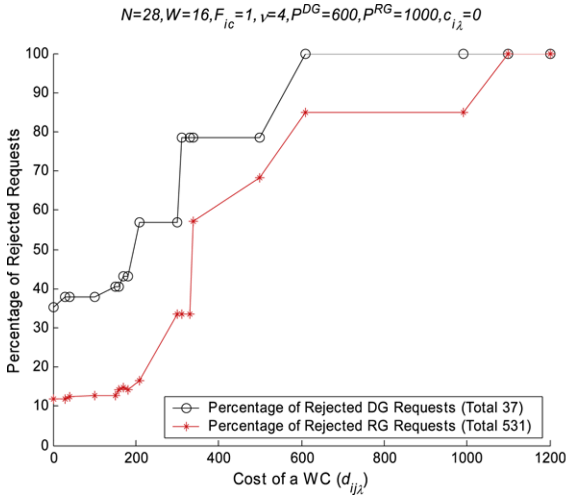

Fig. 11. Percentage of rejected DG and RG requests as the cost of a WC increases.

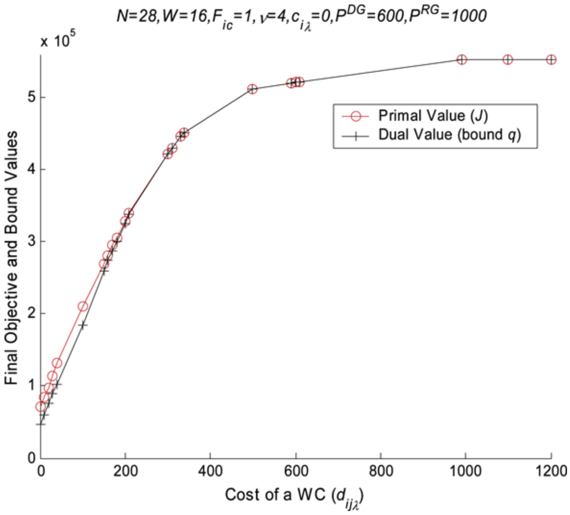

Fig. 12. The achieved value and a bound of the optimization objective function.

achieved value and a bound for the objective function are shown in Fig. 12. The small gap between the achieved value and the bound indicates the efficiency of our algorithm.

8. Conclusions

We proposed GoS differentiated static resource allocation schemes in WDM networks. For a given network topology and resource configuration, our algorithm allocates resources to lightpath requests according to their corresponding service grades. To achieve an optimized resource allocation scheme, we proposed a penalty-based formulation that penalizes both the use of resources and the rejection of requests. Our formulation allows the easy implementation of penalty assignment schemes to address the specific GoS differentiation and fairness requirements. A solution method based on the Lagrangian relaxation and subgradient methods is used to solve our model. The solution method has polynomial computation complexity, and provides performance bounds to evaluate the optimality of results. We investigated two penalty assignment schemes to support GoS differentiation for the requests between the same node pair, and between different node pairs. In addition, we integrated the cost factor into the static GoS differentiation model. We conducted computations for 14-, 22- and 28-node mesh topology networks with various resource configurations and traffic patterns. Our computation results show consistent trends in all computation cases. In this paper, we presented the results of the 28-node network, since the larger a network is, the more challenging it is to optimize the resource allocation. Our results demonstrate the effectiveness of our algorithm and provide a guideline to the network operators to properly devise their own penalty assignment schemes.

Acknowledgements

This research was partially supported by the National Science and Engineering Research Council (NSERC) of Canada, through the Agile All-Photonic Networks (AAPN) Research Network. The authors want to thank Prof. Oliver Yang (University of Ottawa, Canada) for his constructive comments to improve the quality of the paper.

Appendix. A heuristic algorithm to derive a feasible solution to the original problem from the solution to the Lagrangian dual problem

A heuristic algorithm is used to derive a feasible solution to the original problem from the solution to the Lagrangian dual problem. The algorithm decides whether a request should be accepted or rejected, and which RWA scheme should a lightpath use. Our heuristic algorithm is shown in the flowchart in Fig. 13. The following rules are used to determine the priority of requests:

Rule 1. The request with the highest rejection penalty is assigned to the highest priority.

Rule 2. For requests with the same rejection penalty, the one with the lowest hop number in the dual solution is assigned to the highest priority. Further ties are broken randomly.

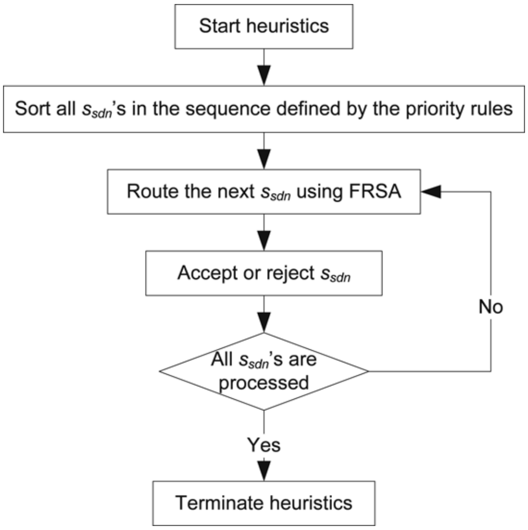

Fig. 13. The flowchart of the heuristic algorithm to obtain a feasible solution.

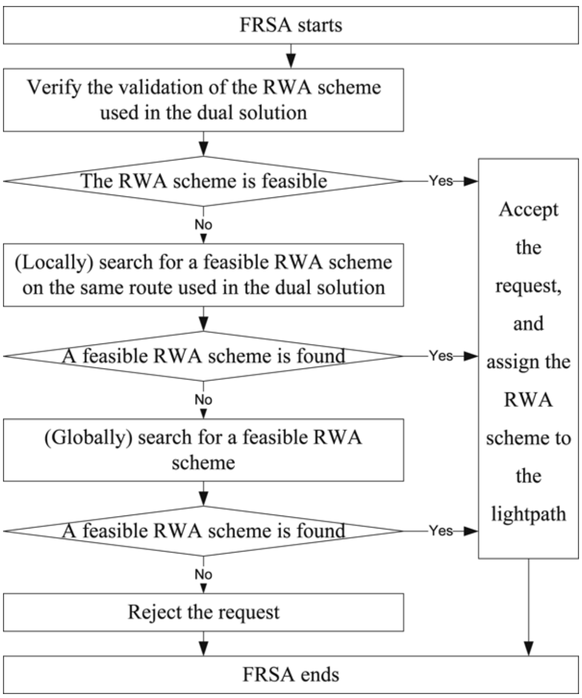

Fig. 14. Flowchart of the feasible route searching algorithm (FRSA).

After determining the priority of requests, we use a Feasible Route Searching Algorithm (FRSA) sequentially for each request. Essentially, FRSA searches for a feasible RWA scheme on a wavelength graph. It uses the guideline of the RWA scheme obtained in the dual solution. The wavelength graph is a directional graph that represents available resources in the network. A flowchart of FRSA is shown in Fig. 14, followed by its detailed description. The computational complexity of the FRSA is $O\left((NW)^2\right)$, since each node is divided into W vertices, and the total number of vertices in a complete wavelength graph is $N \times W$.

To construct a complete wavelength graph, the following procedure is used:

1. Divide every node into W vertices to represent the different wavelengths used by the node;
2. Connect the vertices for the same node (but different wavelengths) with arcs to represent wavelength converters at the node. The weight of these arcs is assigned to $c_{i\lambda}$, which is the cost of a wavelength converter at node i to convert an incoming lightpath with wavelength λ;
3. Connect the vertices for two neighbour nodes (on the same wavelength) with arcs to represent wavelength channels between the two nodes. The weight of these arcs is assigned to d_{ijc}, which is the cost of a wavelength channel from node i to node j on wavelength c.

The Feasible Route Searching Algorithm (FRSA):
Step 1 (Validation of the dual solution):

Verify the validation of the RWA scheme used in the dual solution by checking whether the relaxed two constraints are satisfied. First of all, the RWA scheme must not use any wavelength channel that is already used by other requests with higher priority. Secondly, if the RWA scheme uses a wavelength converter, the same type of wavelength converter must be available after higher priority requests are handled.

(1.1) If the RWA scheme is feasible, the same RWA scheme is assigned to this request. FRSA ends;
(1.2) If the RWA scheme is not feasible, go to *Step* 2.

Step 2 (Search for a feasible RWA scheme on the same route):

Search for a feasible RWA scheme on the same route used in the dual solution. A tailored wavelength graph is composed of only the links and nodes used in the dual solution for this request. The shortest path first algorithm is applied to the tailored wavelength graph.

(2.1) If a feasible RWA scheme is found, the RWA scheme is assigned to this request. FRSA ends;
(2.2) If no feasible RWA scheme is found, go to *Step* 3.

Step 3 (Globally search for a feasible RWA scheme):

A complete wavelength graph is composed of all available wavelength channels and converters. It represents all the resources that are not assigned to higher priority requests. The shortest path first algorithm is applied to this complete wavelength graph.

(3.1) If a feasible RWA scheme is found, the RWA scheme is assigned to this request. FRSA ends;

(3.2) If no feasible RWA scheme is found, reject the request. FRSA ends.

References

[1] G.N. Rouskas, Routing and wavelength assignment in optical WDM networks, in: J. Proakis (Ed.), Wiley Encyclopedia of Telecommunications, John Wiley & Sons Inc., NJ, USA, 2001.

[2] J. Zheng, H.T. Mouftah, Optical WDM Networks: Concepts and Design Principles, Wiley-IEEE Press, Hoboken, NJ, 2004.

[3] A. Jukan, H.R. van As, Service-specific resource allocation in WDM networks with quality constraints, IEEE Journal on Selected Areas in Communications 18 (10) (2000) 2051–2061.

[4] A. Jukan, G. Franzl, Path selection methods with multiple constraints in service-guaranteed WDM networks, IEEE/ACM Transactions on Networking 12 (6) (2004) 59–72.

[5] T. Tachibana, S. Kasahara, QoS-guaranteed wavelength allocation for WDM networks with limited-range wavelength conversion, IEICE Transactions on Communications E87-B (6) (2004) 1439–1450.

[6] K. Mosharaf, J. Talim, I. Lambadaris, Optimal resource allocation and fairness control in all-optical WDM networks, IEEE Journal on Selected Areas in Communications 23 (8) (2005) 1496–1507.

[7] A. Szymanski, A. Lason, J. Rzasa, A. Jajszczyk, Grade-of-service-based routing in optical networks, IEEE Communications Magazine 45 (2) (2007) 82–87.

[8] Y. Chen, M. Hamdi, D.H.K. Tsang, Proportional QoS over WDM networks: Blocking probability, in: Proceedings of the 6th IEEE Symposium on Computers and Communications, ISCC, Hammamet, Tunisia, 2001, pp. 210–215.

[9] Y. Zhang, O. Yang, H. Liu, A Lagrangean relaxation and subgradient framework for the routing and wavelength assignment problem in WDM networks, IEEE Journal on Selected Areas in Communications 22 (9) (2004) 1752–1765.

[10] J. Li, H. Zhang, A new solution to the K-shortest paths problem and its application in wavelength routed optical networks, Photonic Network Communications 12 (3) (2006) 269–284.

[11] J. Yates, J. Lacey, D. Everitt, M. Summerfield, Limited-range wavelength translation in all-optical networks, in: Proceedings of the 15th Annual Joint Conference of the IEEE Computer and Communications Societies, INFOCOM, vol. 3, San Francisco, California, USA, 1996, pp. 954–961.

[12] M.D. Swaminathan, K.N. Sivarajan, Practical routing and wavelength assignment algorithms for all optical networks with limited wavelength conversion, in: Proceedings of IEEE International Conference on Communications, ICC, vol. 5, New York, USA, 2002, pp. 2750–2755.

[13] K.C. Lee, V.O.K. Li, A wavelength-convertible optical network, IEEE/OSA Journal of Lightwave Technology 11 (5–6) (1993) 962–970.

[14] I. Chlamtac, A. Farago, T. Zhang, Lightpath (wavelength) routing in large WDM networks, IEEE Journal on Selected Areas in Communications 14 (5) (1996) 909–913.

[15] D.P. Bertsekas, Nonlinear Programming, second ed., Athena Scientific, Belmont, MA, USA, 1999.

[16] D.J. Hoitomt, P.B. Luh, K.R. Pattipati, A practical approach to job-shop scheduling problems, IEEE Transactions on Robotics and Automation 9 (1) (1993) 1–13.