Optimal Lightpath Establishment Policy for Service Differentiation in WDM Networks with Full-Range Wavelength Conversion

Takuji TACHIBANA and Shoji KASAHARA

Graduate School of Information Science,
Nara Institute of Science and Technology
Takayama 8916-5, Ikoma, Nara 630-0192, Japan
Email: {takuji-t, kasahara}@is.naist.jp
Tel: +81–743–72–5351, Fax: +81–743–72–5359

Abstract

In this paper, we consider a lightpath establishment policy for WDM networks with full-range wavelength conversion capability. The optimal policy considered here is such that the service differentiation for the connection loss probability is provided while the total connection loss probability is small. We derive the optimal policy for the lightpath establishment based on Markov Decision Process (MDP). In numerical examples, we show that the optimal lightpath establishment policy can provide the service differentiation for the connection loss probability, decreasing the total connection loss probability. We also show the effectiveness of the optimum lightpath establishment policy in a uni-directional ring network.

Index Terms: Service differentiation, Markov decision process, Lightpath establishment policy, Connection loss probability, WDM networks

I. Introduction

In wavelength routing networks, a connection called *lightpath* is established with wavelength between end nodes and data is transmitted over the lightpath [10]. Because the lightpath is routed optically from one link to another link at each node, the data transmission does not require opto-electronic-optic (O/E/O) conversion.

In dynamic lightpath configuration, when a lightpath-establishment request arrives at a node, an available wavelength is allocated to the lightpath. If all the wavelength allocations in the nodes along the path succeed, the lightpath is eventually established. When there is no data transmission in the established lightpath, the lightpath is released. If there is no available wavelength in some node along the path, the lightpath establishment fails.

The recent increase of Internet users and the diversity of network applications have required the service differentiation in WDM networks. In the wavelength routing networks with the dynamic lightpath configuration, the service differentiation for the connection loss probability has been studied [3], [4], [8], [9]. When the networks have the full-range wavelength capability, the number of available wavelengths for each node is pre-determined so that higher priority class can utilize more wavelengths than lower classes [3], [11]. This can simply provide the service differentiation for the connection loss probability, however, wavelengths are not utilized efficiently due to the static wavelength allocation, and this results in small wavelength utilization and high total loss probability.

In this paper, we consider a lightpath establishment policy for the wavelength routing networks with the dynamic lightpath configuration. In order both to provide the service differentiation and to decrease the total loss probability, we derive the optimal lightpath establishment policy using a Markov Decision Process (MDP).

[5] studied the optimal wavelength allocation policy for a two-hop WDM network using the MDP. In this MDP model, the number of traffic classes is limited, and a decision for the wavelength allocation policy is made at each node only when a lightpath departs from the network.

In this paper, we further develop the model studied in [5] so that a policy decision is made at both the arrival and departure epochs of a lightpath. In addition, multiple priority classes are provided. In the developed model, the event of connection loss as well as the number of used wavelengths is taken into consideration, and the reward function is defined as the sum of reward for the lightpath establishment success and cost for the failure of lightpath establishment. With the resulting optimal policy, we evaluate by simulation the connection loss probability for a uni-directional ring network, and show the effectiveness of the optimal lightpath establishment policy. Moreover, we investigate how the reward function affects the lightpath establishment.

The rest of the paper is organized as follows. Section II summarizes the static wavelength allocation for the service differentiation of the connection loss probability. In Section III, we present the lightpath establishment control for the service differentiation. In Section IV, we formulate the method as the MDP and derive the optimal lightpath establishment policy. Numerical examples are shown in Section V and conclusions are presented in Section VI.

II. Static Wavelength Allocation for Service Differentiation

In this section, we summarize the static wavelength allocation for the service differentiation [3], [11] which is considered as a conventional method in the remainder of the paper.

We consider a wavelength routing network with the capability of the full-range wavelength conversion. Let W denote the number of wavelengths multiplexed into an optical fiber. In this wavelength routing network, M classes require different acceptable loss probabilities. The M classes are numbered from 1 to M and the class i has high priority over the class j when $i < j$, and the class i requires smaller connection loss probability than the class j. Therefore, connections of the class 1 have the highest priority and requires the smallest loss probability.

In order to provide multiple service classes for the connection loss probability, the number of available wavelengths for each service class is pre-determined. The number of available wavelengths for the class i $(i = 1, \cdots, M)$ is W_i, and W_i satisfies the following.

$$0 \leq W_M < \cdots < W_i < \cdots < W_1 = W. \tag{1}$$

The inequalities (1) imply that higher priority class can use more wavelengths and it is expected that the resulting connection loss probability of high priority class becomes small. Therefore, the service differentiation is provided in terms of the connection loss probability.

III. Lightpath Establishment Control for Service Differentiation

In this section, we consider a lightpath establishment control for the service differentiation. Each node has a controller for the lightpath establishment and the controller investigates the number of established lightpaths for each class at an output link. This investigation is performed just after either a connection arrival or a connection departure. According to the state of the link, the controller decides which class is accepted for the next lightpath establishment. In other words, the controller decides an action for the next lightpath arrival among an action set, say $\mathbf{A}$. The set $\mathbf{A}$ includes the following actions.

- A lightpath for the class i $(i = 1, \cdots, M)$ can be established.
- A lightpath for the two classes i and j $(i \neq j)$ can be established.
- A lightpath for the m $(2 < m < M)$ classes $i_1, i_2, \cdots, i_m$ can be established.
- A lightpath for all the M classes can be established,

Note that, when the number of service classes is M, the number of possible actions is $2^M - 1$.

Let N_i denote the number of the established lightpaths for the class i $(i = 1, \cdots, M)$. Note that $0 \leq N_i \leq W$ and $\sum_{i=1}^{M} N_i \leq W$. We define the current state of the link, s, as

$$s = (N_1, N_2, \cdots, N_M, I), \tag{2}$$

where I is given by

$$I = \begin{cases} 1, & \text{if a connection loss occurred at the last event,} \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

Let $\mathbf{S}$ denote the system state space. The event considered here is either a connection arrival or a connection departure, and hence I takes 0 if the last event is a connection departure. The controller decides an action a among the set $\mathbf{A}$ according to the current state s $(\in \mathbf{S})$. This mapping between state s $(\in \mathbf{S})$ and action a $(\in \mathbf{A})$ is denoted as a policy π, and π indicates which action should be taken in each state.

For example, suppose that some policy π is adopted. According to the policy π, any state has the corresponding action. When a connection request for the class i arrives at the node and the output link of the node is in some state s $(\in \mathbf{S})$, the controller decides whether the lightpath is established or not according to the current action a $(\in \mathbf{A})$. If the state of the link s changes to a state s' $(\in \mathbf{S})$ according to the action a, the controller decides a new action a' $(\in \mathbf{A})$ based on the state s'. When a connection is released, the state of the link changes to a new state s'' $(\in \mathbf{S})$ and an action a'' $(\in \mathbf{A})$ is decided based on the state s''.

The policy π significantly affects the effectiveness of this lightpath establishment control. In the next section, we develop an MDP model for a single link network and derive the optimal lightpath establishment policy. The readers are referred to [2], [6], [7] for the detail of the MDP.

IV. Optimal Policy based on Markov Decision Process

We focus on a single link in the wavelength routing network and use the following assumptions.

- W wavelengths are multiplexed at the single link.
- The number of classes is three and the class i $(i = 1, 2, 3)$ has priority over the class j if $i < j$.
- Connections of the class i arrive at the single link according to a Poisson process with rate λ_i.
- Connection holding time of the class i is exponentially distributed with rate μ_i.
- No queueing for connection request is permitted, that is, the connection is lost immediately after the connection establishment fails.

Let s_t denote the system state at time t. From the above assumptions, the process $\{s_t; t \geq 0\}$ is a continuous-time Markov chain whose transitions are either the event of an arrival or a departure of a connection.

In this section, first, we formulate this lightpath establishment problem as a continuous-time MDP. Next, we translate the initial continuous-time MDP into an equivalent discrete-time MDP. Finally, we derive the optimal policy for the lightpath establishment with the discrete-time MDP.

TABLE I
STATE TRANSITION IN MDP.

Current state $s=(N_1,N_2,N_3,I)$	Event	Action a	Next state s'	Transition probability $P^a_{ss'}$
$K<W$	Arrival of class 1	1,4,6,7	$(N_1+1,N_2,N_3,0)$	λ_1/ν
		2,3,5	$(N_1,N_2,N_3,1)$	λ_1/ν
	Arrival of class 2	2,4,5,7	$(N_1,N_2+1,N_3,0)$	λ_2/ν
		1,3,6	$(N_1,N_2,N_3,1)$	λ_2/ν
	Arrival of class 3	3,5,6,7	$(N_1,N_2,N_3+1,0)$	λ_3/ν
		1,2,4	$(N_1,N_2,N_3,1)$	λ_3/ν
$K=W$	Arrival of any class	1,2,3,4,5,6,7	$(N_1,N_2,N_3,1)$	λ_{all}/ν
$N_1>0$	Departure of class 1	1,2,3,4,5,6,7	$(N_1-1,N_2,N_3,0)$	μ_1/ν
$N_2>0$	Departure of class 2	1,2,3,4,5,6,7	$(N_1,N_2-1,N_3,0)$	μ_2/ν
$N_3>0$	Departure of class 3	1,2,3,4,5,6,7	$(N_1,N_2,N_3-1,0)$	μ_3/ν

A. Problem Formulation

The MDP consists of an action space, a state space, a state-transition structure, and a reward structure [1], [12], and we explain those in the following.

Action space: Let $\mathbf{A}$ denote the action set, and an action a ($a=1,\cdots,7$) in the set is shown as follows.

1. A lightpath of the class 1 can be established.
2. A lightpath of the class 2 can be established.
3. A lightpath of the class 3 can be established.
4. A lightpath for the classes 1 and 2 can be established.
5. A lightpath for the classes 2 and 3 can be established.
6. A lightpath for the classes 3 and 1 can be established.
7. A lightpath for the classes 1, 2, and 3 can be established.

The controller decides an action a among the set $\mathbf{A}$ for the next arrival or departure of a lightpath after each event.

State space: The state space of this system is $\mathbf{S}$ and a state of the system is denoted by s ($\in \mathbf{S}$). As shown in the previous section, the system state s is given by the vector (N_1,N_2,N_3,I) where $0 \le N_i \le W$ and $I=0,1$. Note that $\sum_{i=1}^{3} N_i \le W$.

State transition: Let $K=N_1+N_2+N_3$ denote the total number of the established lightpath. Assume that the current state is $s=(N_1,N_2,N_3,I)$. We show possible state transitions from s in Table I. Table I also shows $P^a_{ss'}$ which will be explained in the following subsection.

Reward Structure: We consider the transition from the state $s=(N_1,N_2,N_3,I)$ to the state $s'=(N'_1,N'_2,N'_3,I')$ with some action a. Let r_i ($i=1,2,3$) denote the reward assigned to the lightpath establishment for the class i. The rewards satisfy the following inequalities in order that high priority class can establish more lightpaths than low priority class,

$$0 < r_3 < r_2 < r_1 \le 1. \tag{4}$$

We define the one-step reward function $R^a_{ss'}$ associated with the transition from s to s' under the action a as

$$R^a_{ss'} = \alpha(r_1 N'_1 + r_2 N'_2 + r_3 N'_3) - (1-\alpha) I'. \tag{5}$$

where $0 \leq \alpha \leq 1$ is the weighted parameter for the service differentiation and the connection loss.

$R^a_{ss'}$ depends only on the state s' and does not depend directly on the state s and the action a. However, note that s' depends on s and a through the above state transition. As α becomes small, the connection loss becomes more important factor than the service differentiation.

Optimal goal: The optimal goal is to determine the lightpath establishment policy which maximizes the expected reward over infinite time.

B. Uniformization of the Continuous-Time MDP

In the previous subsection, we formulated the lightpath establishment problem as a continuous-time MDP. In this subsection, we convert the continuous-time MDP into an equivalent discrete-time MDP by the well-known uniformization technique in order to utilize dynamic programming [2].

For the uniformization, we introduce a random sampling rate $\nu = \lambda_1 + \lambda_2 + \lambda_3 + W(\mu_1 + \mu_2 + \mu_3)$. With this random sampling rate, we can convert the continuous-time MDP into an equivalent discrete-time MDP. In the discrete-time MDP, only one single transition can occur during each time slot, and in addition to the arrival and departure of lightpaths, fictitious event with which the system state does not change occurs. Let $P^a_{ss'}$ denote the state transition probability from s to s' under the action a. We show $P^a_{ss'}$ in Table I, except for the fictitious event. In Table I, $\lambda_{all} = \lambda_1 + \lambda_2 + \lambda_3$. For the fictitious event, we have $P^a_{ss} = \{(W - N_1)\mu_1 + (W - N_2)\mu_2 + (W - N_3)\mu_3\}/\nu$.

C. Derivation of Optimal Policy

For the discrete-time MDP obtained in the previous subsection, a policy π of the MDP is a mapping between states s and actions a, and it indicates which action should be taken in each state. Under the policy π, let $V^\pi(s)$ denote the expected discounted reward. For each state s, $V^\pi(s)$ is given by

$$V^\pi(s) = \sum_{s'} P^a_{ss'}[R^a_{ss'} + \gamma V^\pi_n(s')], \tag{6}$$

Our objective is to obtain the stationary policy π^* that maximizes the discounted reward $V^\pi(s)$.

Let $V_n(s)$ denote the expected discounted reward over n time steps from the initial state s. Since the state space $\mathbf{S}$ is finite and both the state transition probability and the one-step reward are explicitly given, the following iteration algorithm

$$V_{n+1}(s) = \max_a \sum_{s'} P^a_{ss'}[R^a_{ss'} + \gamma V_n(s')] \tag{7}$$

converges to the optimal value function [2], [5], [7]. Therefore, $V^{\pi^*}(s)$ satisfies the following equation

$$V^{\pi^*}(s) = \lim_{n\to\infty} V_n(s), \tag{8}$$

and we can obtain the optimal policy π^* numerically.

V. Numerical Examples

A. Optimal Policy

In this subsection, we investigate the characteristics of the optimal policy for the service differentiation.

TABLE II

OPTIMAL POLICY IN THE CASE OF $\lambda_1 = \lambda_2 = \lambda_3 = 3.0$.

(a) Impact of the reward r_2 ($r_1 = 1.0, r_3 = 0.1, \alpha = 1.0$).

State	$(0,0,0)$	$(7,0,0)$	$(2,2,2)$	$(0,7,0)$	$(0,0,7)$
$r_2 = 0.9$	7	4	4	4	4
$r_2 = 0.4$	7	1	4	1	1
$r_2 = 0.2$	7	1	1	1	1

(b) Impact of the weighted parameter α ($r_1 = 1.0, r_2 = 0.4, r_3 = 0.1$).

State	$(0,0,0)$	$(7,0,0)$	$(2,2,2)$	$(0,7,0)$	$(0,0,7)$
$\alpha = 0.6$	7	1	4	1	1
$\alpha = 0.3$	7	4	4	4	4
$\alpha = 0.01$	7	7	7	7	7

A.1 Impacts of the reward and weighted parameter

First, we investigate how the reward r_2 for the class 2 and the weighted parameter α affect the optimal policy. In the following, we set $W = 8$, $\lambda_1 = \lambda_2 = \lambda_3 = 3.0$ and $\mu_1 = \mu_2 = \mu_3 = 1.0$.

Table II (a) shows the optimal actions for some states against r_2 in the case of $r_1 = 1.0$, $r_3 = 0.1$, and $\alpha = 1.0$ [1]. We obtained the result that for any N_1, N_2 and N_3, the optimal action $(N_1, N_2, N_3, 0)$ is the same as $(N_1, N_2, N_3, 1)$. Therefore the value of I is omitted from the state in Table II.

From the case of $r_2 = 0.9$ in this table, we observe that the action 4 where the classes 1 and 2 can establish a lightpath is optimal, except for $(0, 0, 0)$. This is because the rewards for the classes 1 and 2 are larger than that for the class 3. When r_2 becomes small, the priority of the class 2 also becomes small. As a result, the action 1 where the class 1 can establish a lightpath becomes optimal instead of the action 4. We verified that several types of the service differentiation can be provided by changing the rewards [2].

On the other hand, Table II (b) shows the optimal actions for some states against α in the case of $r_1 = 1.0$, $r_2 = 0.4$, and $r_3 = 0.1$. From the case of $\alpha = 0.6$ in this table, we observe that three actions are obtained for the optimal policy, and this implies that the service differentiation can be provided with those actions. When α becomes large, the weight of the connection loss in the reward function (5) becomes large. As a result, the action 7 with which a connection request for all classes is accepted becomes optimal for any states. This implies that the small α cannot provide the service differentiation.

A.2 Impact of the connection holding time of each class

Next, we investigate how the holding time of each class affects the optimal policy. Here, we set $W = 8$, $\lambda_1 = \lambda_2 = \lambda_3 = 1.0$, $r_1 = 1.0$, $r_2 = 0.4$, and $r_3 = 0.1$.

Tables III (a) and (b) show the optimal actions for some states against both $1/\mu_1$ and $1/\mu_2$ in the cases of $\alpha = 0.6$ and $\alpha = 0.2$, respectively. In both tables, we set $1/\mu_3 = 1.0$. From the case of $1/\mu_1 = 1/\mu_2 = 1/\mu_3 = 1.0$ (the first row in Table III (a)), we find that the service differentiation can be provided with the action 4. When $1/\mu_1$ and $1/\mu_2$ become large, the action 1 where only the class 1 can establish a lightpath becomes optimal. This results from the large reward r_1 even though the holding time of the class 1 is the same as that of the class 2.

[1] The optimal actions for other states are omitted due to the page limitation.

[2] We investigated the effectiveness of r_1 and r_3 and obtained the results similar to r_2.

TABLE III

OPTIMAL POLICY IN THE CASE OF $r_1 = 1.0, r_2 = 0.4,$ AND $r_3 = 0.1$.

(a) Impacts of the holding times $1/\mu_1$ and $1/\mu_2$ ($\alpha = 0.6$).

State	$(0,0,0)$	$(7,0,0)$	$(2,2,2)$	$(0,7,0)$	$(0,0,7)$
$1/\mu_{1(2)} = 10^0$	7	4	7	4	4
$1/\mu_{1(2)} = 10^1$	7	1	1	1	1
$1/\mu_{1(2)} = 10^3$	6	1	1	1	1

(b) Impacts of the holding times $1/\mu_1$ and $1/\mu_2$ ($\alpha = 0.2$).

State	$(0,0,0)$	$(7,0,0)$	$(2,2,2)$	$(0,7,0)$	$(0,0,7)$
$1/\mu_{1(2)} = 10^0$	7	7	7	7	7
$1/\mu_{1(2)} = 10^1$	7	6	6	6	1
$1/\mu_{1(2)} = 10^3$	6	3	6	3	1

On the other hand, in the case of $\alpha = 0.2$ (see Table III (b)), the weight of the connection loss is larger than that of the service differentiation. Therefore, when $1/\mu_1$ and $1/\mu_2$ are equal to 1.0, the action 7 where all classes can establish a lightpath is optimal. As $1/\mu_1$ and $1/\mu_2$ become large, the actions 3 and 6 become optimal. Note that in both actions, the class 3 which is the lowest priority class can establish a lightpath. This result implies that the optimal policy considerably depends on the holding time of high priority class when the connection loss is more important factor in the policy than the service differentiation.

B. Performance Comparison

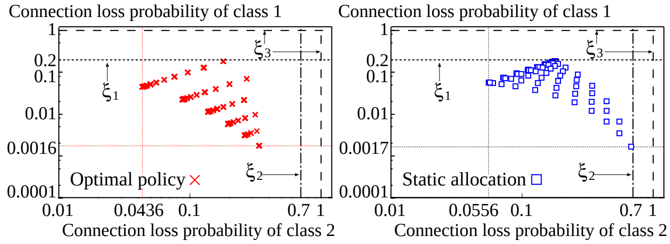
In this subsection, we compare the performance of the optimal lightpath establishment policy with that of the static wavelength allocation in Sect. II. We consider a uni-directional ring network with 10 nodes which have full-range wavelength conversion capability. We set $W = 16$, $\lambda_1 = \lambda_2 = \lambda_3 = 10$, and $\mu_1 = \mu_2 = \mu_3 = 1.0$. The pair of source and destination nodes of a connection is distributed uniformly, i.e., any pair is selected with the same probability.

Here, we change the rewards r_1, r_2, and r_3 from 0.05 to 1.0 with 0.05 increment, respectively, and α is set to 0.1, 0.5, 0.8, and 1.0. For each parameter set (r_1, r_2, r_3, α), we firstly derived the optimal lightpath establishment policy with the MDP. Then, with the resulting optimal policy for the parameter set, the connection loss probability $P_{loss}^{(i)}$ $(i = 1, 2, 3)$ of the class i in the ring network was calculated by simulation. We assume that each class requires that the connection loss probability $P_{loss}^{(i)}$ is smaller than or equal to the constant ξ_i.

Figs. 1 (a) and (b) show the results for the optimal poloicy and the static wavelength allocation, respectively, in the case of $\xi_1 = 0.2$, $\xi_2 = 0.7$, and $\xi_3 = 1.0$. The loss probabilities of the classes 1 and 2 such that $P_{loss}^{(1)} \le \xi_1$, $P_{loss}^{(2)} \le \xi_2$, and $P_{loss}^{(3)} \le \xi_3$ are plotted in Fig. 1.

From Fig. 1 (a), we observe that the optimal lightpath establishment policy can provide several types of the service differentiation in terms of the connection loss probability. We also find that the loss probabilities of the classes 1 and 2 for the optimal policy are smaller than those for the static wavelength allocation. This implies that wavelengths are utilized effectively with the optimal policy in the case of $\xi_1 = 0.2$, $\xi_2 = 0.7$, and $\xi_3 = 1.0$.

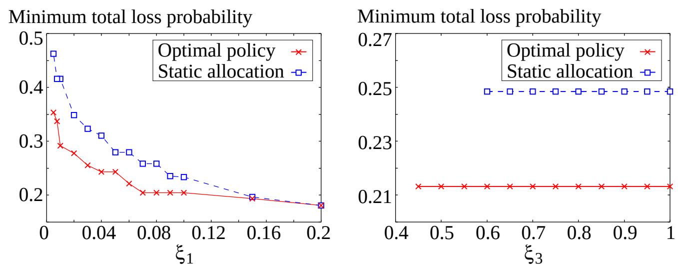
Next, we consider the case where each class requires smaller loss probability than the Fig. 1 case. Fig. 2(a) ((b)) shows how ξ_1 (ξ_3) affects the minimum total loss probabilities for both methods. Each total loss probability plotted in both figures is the minimum among the set where $P_{loss}^{(i)} \le \xi_i$ for $i = 1, 2,$ and 3.



(a) Optimal policy.

(b) Static allocation.

Fig. 1. Performance comparison of the optimal policy and the static wavelength allocation in the case of $\xi_1 = 0.2, \xi_2 = 0.7$, and $\xi_3 = 1.0$.



(a) Impact of ξ_1 ($\xi_2 = 0.7$ and $\xi_3 = 1.0$).

(b) Impact of ξ_3 ($\xi_1 = 0.1$ and $\xi_2 = 0.15$).

Fig. 2. Minimum total loss probability vs. ξ_1 and ξ_3.

First, we consider the case where the class 1 requires a loss probability smaller than or equal to 0.2. From Fig. 2(a), we observe that the minimum total loss probabilities for both methods are almost the same in the case of $\xi_1 = 0.2$. When the class 1 requires small loss probability, i.e., ξ_1 becomes small, the minimum total loss probability for the optimal policy becomes smaller than that for the static wavelength allocation. This is because the optimal policy is determined so as to decrease the total loss probability with small α. Therefore, the optimal lightpath establishment policy can use wavelengths effectively when high priority class requires a small loss probability.

Next, we consider the case where the class 3 requires a loss probability smaller than or equal to 1.0 when $\xi_1 = 0.1$ and $\xi_2 = 0.15$. From Fig. 2(b), we observe that the minimum total loss probability for the optimal policy is smaller than that for the static wavelength allocation regardless of ξ_3. Moreover, the optimal policy can provide the service differentiation even when ξ_3 is smaller than or equal to 0.6. In this case, the static wavelength allocation cannot provide the service differentiation. Hence, the optimal lightpath establishment policy can also use wavelengths effectively when the low priority class requires small loss probability.

VI. Conclusions

In this paper, we considered the optimal lightpath establishment policy for the service differentiation in wavelength routing networks with full-range wavelength conversion. In numerical examples, we found that the optimal lightpath establishment policy is determined depending on the reward for each class, the weighted parameter, and the lightpath holding time. Simulation results showed that the optimal policy can provide the service differentiation, decreasing the total connection loss probability in the uni-directional ring network. We also observed that the optimal lightpath establishment policy is effective when each class requires a small loss probability. Currently, the study of the optimal lightpath establishment policy in wavelength routing networks with limited-range wavelength conversion is in progress.

References

[1] I. Baldine and G. N. Rouskas, "Dynamic Reconfiguration Policies for WDM Networks," in *Proc. IEEE INFOCOM'99,* Mar. 1999, pp. 313-320.

[2] D. P. Bertsekas, *Dynamic Programming Deterministic and Stochastic Models.* Prentice-Hall, Englewood cliffs, 1987.

[3] A. Jukan, *QoS-based Wavelength Routing in Multi-Service WDM Networks.* New York: Springer, 2000.

[4] A. Jukan and H. R. van As, "Service-Specific Resource Allocation in WDM Networks with Quality Constraints," *IEEE J. Select. Areas Commun.,* vol. 18, no. 10, pp. 2051-2061, Oct. 2000.

[5] K. Mosharaf, J. Talim, and I. Lambadaris, "A Markov Decision Process Model for Dynamic Wavelength Allocation in WDM Networks," in *Proc. Globecom'03*, Dec. 2003.

[6] M. L. Puterman, *Markov Decision Process.* New York: Wiley Inter-Science, 1994.

[7] R. S. Sutton and A. G. Barto, *Reinforcement Learning.* Cambridge: MIT Press, 1998.

[8] T. Tachibana and S. Kasahara, "QoS-Guaranteed Wavelength Allocation for WDM Networks with Limited-Range Wavelength Conversion," *IEICE Trans. Commun.*, vol. E87–B, no. 6, pp. 1439–1450, June 2004.

[9] T. Tachibana and S. Kasahara, "QoS Provisioning with Shared Wavelength Allocation for Limited-range Wavelength Conversion," To appear. (http://genesis.aist-nara.ac.jp/output/paper/04/onm_special_issue.pdf)

[10] R. Ramaswami and K. N. Sivarajan, *Optical Networks: A Practical Perspective.* San Francisco: Morgan Kaufmann Publishers, 1998.

[11] K. W. Ross, and D. H. K. Tsang, "The Stochastic Knapsack Problem," *IEEE Trans. on Commun.*, vol. 37, no. 7, pp. 740-747, July 1989.

[12] G. Wu, E. K. P. Chong, and R. Givan, "Congestion Control via Online Sampling," in *Proc. IEEE INFOCOM'01,* Apr. 2001, pp. 1271-1280.