A Markov Decision Process Model for Dynamic Wavelength Allocation in WDM Networks

Kayvan Mosharaf, Jérôme Talim and Ioannis Lambadaris
Department of Systems and Computer Engineering
Carleton University, 1125 Colonel By Drive
Ottawa, Ontario, Canada, K1S 5B6
Emails: {mosharaf, jtalim, ioannis}@sce.carleton.ca

***Abstract*— This paper[1] outlines an optimal dynamic wavelength allocation in all-optical WDM networks. A simple topology consists of a 2-hop path network with three nodes is studied for three classes of traffic where each class corresponds to different source-destination pair. For each class, call interarrival and holding times are exponentially distributed. The objective is to determine a wavelength allocation policy in order to maximize the weighted sum of users of all classes. Consequently, this method is able to provide differentiated services in the network. The problem can be formulated as a Markov Decision Process to compute the optimal resource allocation policy. It has been shown numerically that for two and three classes of users, the optimal policy is of threshold type and monotonic. Simulation results compare the performance of the optimal policy, with that of Complete Sharing and Complete Partitioning policies.**

I. Introduction

Wavelength Division Multiplexing (WDM), using wavelength routing, is one of the candidates to handle the bandwidth of future wide area backbone networks. In wavelength routing networks, each optical path must be established with a specific wavelength between each source-destination pair. This is known as wavelength continuity constraint and can be relaxed by using wavelength converters at intermediate nodes [1]. The routing and wavelength assignment problem (RWA) is an important issue in WDM networks. RWA is usually divided into two separate sub-problems: i) wavelength assignment problem and ii) routing problem. Many heuristic algorithms such as random wavelength assignment and first-fit have been already proposed [2]. The objective of these algorithms is typically to minimize the overall call blocking probability or maximize the overall utilization in a single-class network. Few consistent results dealing with service differentiation in all-optical networks are available. One can refer to [3] for a general analysis of this problem.

In this paper, we investigate the wavelength allocation problem for different classes of users with different priorities. With the objective to maximize the weighted sum of class-based utilization, we define a Markov Decision Process (MDP) problem [4], based on which the optimal wavelength allocation policy can be determined. In many admission control and resource allocation problems in telecommunications [5], [6], it was shown that under some conditions, the optimal policy of an MDP exists and it is stationary and monotone. The Policy Iteration algorithm can be deployed to determine the optimal policy [4].

The rest of the paper is organized as follows: In Section II, we formulate the problem as an MDP for a simple network topology. Section III deals with the definition of the discounted cost function associated with the problem in the infinite horizon case. Section IV shows the structure of optimal policy and section V compares the performance of the proposed policy with other standard policies. Conclusions are presented in Section VI.

II. Problem Formulation

Consider a 2-hop path network topology for a single fiber circuit-switched wavelength routing network [7] as depicted in Fig. 1. The total number of available wavelengths in the system is W for each hop. Traffic is divided into 3 classes: each class corresponds to different source-destination pair. Class 1 (respectively, Class 3) consists of the users that use hop h_1 (respectively, h_2); Class 2 includes the customers that use both hops h_1, h_2. In this paper, we assume that there is a wavelength converter in node 2. Therefore, a Class 2 call is accepted whenever there is one available wavelength in both hops. Any arriving call is blocked when all wavelengths along its path are used. Blocked calls do not interfere with the system.

To improve the system performance it would be essential to assign a certain number of wavelengths to each class as a function of the current number of customers from different classes. Wavelength allocation policies are a particular problem related to resources allocation policies, such as Complete Sharing (CS) and Complete Partitioning (CP) [8]. When implementing CS, no wavelength is reserved for any class. In addition, an arriving call will be accepted if at least one wavelength is available throughout all the hops along its path.

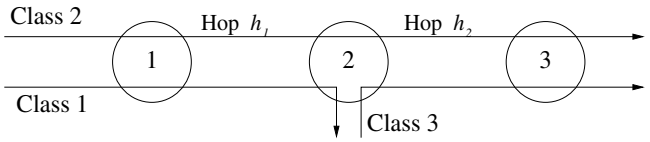

Fig. 1. 2-hop network topology.

[1]This work was supported by a grant from MITACS (Mathematics for Information Technology and Complex Systems)

When deploying CP policy, each class is dedicated a constant number of wavelengths that cannot be used by calls from other classes. This paper investigates dynamic wavelength allocation policies, that we refer to as Dynamic Partitioning (DP). It consists of determining the appropriate number of wavelengths allocated to each class taking into account the current state of the system.

For each class, call interarrival times and holding times are exponentially distributed. Let us define the following notations:

- $1/\lambda_i$, $i = 1, 2, 3$, is the mean interarrival time of Class i calls.
- $1/\mu_i$, $i = 1, 2, 3$, is the mean holding time of Class i calls.
- n_i , $i = 1, 2, 3$, is the number of Class i calls currently in the system.

We will consider first a simpler model which consists of only Class 1 and Class 2 calls.

Let k be the number of wavelengths allocated to Class 2 calls. For any (n_1, n_2), one can derive $i = W - k - n_1$ and $j = k - n_2$ as the numbers of available wavelengths reserved to Class 1 and Class 2, respectively. Therefore, the three-component vector (i, j, k) characterizes completely the system. Let $S = \{s = (i, j, k) | 0 \leq i \leq W - k, 0 \leq j \leq k, 0 \leq k \leq W\} \subset \{0, 1, 2, ...W\}^3$ be the system state space, and s_t be the system state at time t. Based on the statistical assumptions, $\{s_t, t \geq 0\}$ is a continuous-time Markov chain whose transitions are either the event of an arrival or a departure of a call. To simplify the notation, the following operators are introduced:

- A_i, $i = 1, 2$: Arrival operator describing the change of system state at the arrival time of a Class i user.
  - $A_1 s = ((i-1)^+, j, k)$
  - $A_2 s = (i, (j-1)^+, k)$
- D_i, $i = 1, 2$: Departure operator describing the change of system state at the departure time of a Class i user.
  - $D_1 s = (i+1, j, k)$
  - $D_2 s = (i, j+1, k)$

where $x^+ = \max(0, x)$ and $s = (i, j, k)$.

Investigating DP policy involves the calculation of wavelength allocation as a function of the current system $s = (i, j, k)$ and the event e. The objective, then , is to maximize the usage of the optical resources. This equivalently translates into maximizing weighted sum of number of requests of the two classes. This problem can be formulated as a Markov Decision Process (MDP) [9]. Let us describe more accurately the model:

- Decision epochs take place only at departure times, when a call terminates and a wavelength becomes available. This wavelength may be reserved for the same class or switched to the other one. Fig. 2(a) illustrates the initial situation of a system in state $s = (i, j, k)$. Fig. 2(b) provides the final system state, after the departure of a Class 1 call when we decide to keep the wavelength to Class 1. Similarly, Fig. 2(c) shows the system state after the departure of a Class 1 call when the policy maker decides to reserve the wavelength for Class 2. The decision results in increasing the value of k by 1, or leaving it unmodified. Similarly, when considering the termination of a Class 2 call, the decision will result in decreasing the value of k by 1, or leaving it unchanged.
- Therefore, when the system is in state s and the event e has just occurred, the set of possible actions $A(s, e)$ are:

$$A(s,e) = \begin{cases} \{0\} & \text{if } e = A_1 \text{ or } A_2 \\ \{0, +1\} & \text{if } e = D_1 \\ \{-1, 0\} & \text{if } e = D_2 \end{cases}$$

Let P_a, $a = -1, 0, 1$ be the policy operator to describe the change of system state when applying action $a \in A(s, e)$:

- $P_a s = ((i-1)^+, j+1, k+1)$ for $a = 1$
- $P_a s = (i, j, k)$ for $a = 0$,
- $P_a s = (i+1, (j-1)^+, (k-1)^+)$ for $a = -1$

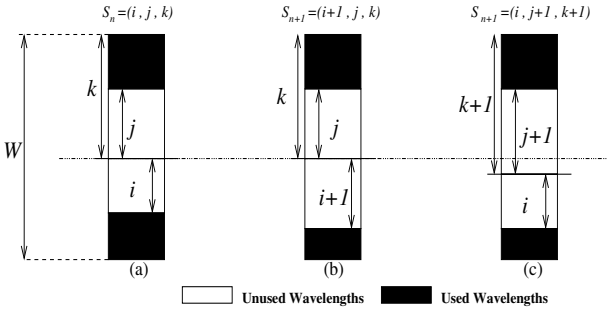

Fig. 2. Possible states after the departure of a Class 1 Call.

This initial continuous-time MDP can be converted into an equivalent discrete-time MDP by applying the uniformization technique [9]. In order to do so, we introduce a random sampling time ν defined as $\nu := W(\mu_1 + \mu_2) + \lambda_1 + \lambda_2$. When considering the equivalent discrete-time MDP, only one single transition can occur during each time slot. And a transition can correspond to an event of: 1) Class 1 call arrival or departure, 2) Class 2 call arrival or departure and 3) fictitious event.

III. The Discounted Cost Problem

Our goal is to determine a wavelength allocation policy that maximizes the weighted sum of n_1 and n_2. To this end, we are going to use the results from MDP for discounted cost model with the infinite horizon [9]. Let us first define the one-step reward:

$$R(s_n) = n_1 + \beta n_2 = W - (i + j\beta + (1-\beta)k) \quad (1)$$

where $s_n = (i, j, k)$ and $\beta \in [0, 1]$ is the weight assigned to Class 2 users. One can notice that maximizing (1) consists of minimizing $(i + j\beta + (1-\beta)k)$. Thus, we can define the one-step cost to be minimized:

$$C(s_n) = B \cdot s_n \quad (2)$$

where vector $B := (1, \beta, (1-\beta))^T$.

The optimal discounted cost function and the optimal policy can be computed by using the following recursive scheme, known as the relative value iteration algorithm [4].

$$V_{n+1}(s) = \min_{\pi}[C(s) + \gamma \sum_{s'} P^{\pi}_{ss'} V_n(s')] \qquad (3)$$

where $P^{\pi}_{ss'}$ is the transition probability to jump from state s to state s' when applying policy π. $P^{\pi}_{ss'}$ is given by:

$$P^{\pi}_{ss'} = \begin{cases} \lambda_1 & \text{if } s' = A_1 s \ , \ a = 0 \\ \lambda_2 & \text{if } s' = A_2 s \ , \ a = 0 \\ n_1\mu_1 & \text{if } s' = D_1 P_0 s, \ a = 0 \\ n_1\mu_1 & \text{if } s' = D_1 P_1 s, \ a = 1 \\ n_2\mu_2 & \text{if } s' = D_2 P_0 s, \ a = 0 \\ n_2\mu_2 & \text{if } s' = D_2 P_1 s, \ a = 1 \\ F & \text{if } s' = s, \ a = 0 \end{cases}$$

where $F = W(\mu_2 + \mu_1) - n_1\mu_1 - n_2\mu_2$

Replacing $P^{\pi}_{ss'}$ in (3) yields:

$$\begin{aligned} V_{n+1}(s) = C(s) + \gamma[&\lambda_1 V_n(A_1 s) + \lambda_2 V_n(A_2 s) + \\ &(W(\mu_2 + \mu_1) - n_2\mu_2 - n_1\mu_1) V_n(s) + \\ &\mu_1 n_1 \ \min\{V_n(D_1 P_0 s), V_n(D_1 P_1 s)\} + \\ &\mu_2 n_2 \ \min\{V_n(D_2 P_0 s), V_n(D_2 P_{-1} s)\}] \end{aligned} \qquad (4)$$

From the above equation, it can be noticed that at a Class 1 call termination time, the optimal action is $a = 0$ if $V_n(D_1 P_0 s) < V_n(D_1 P_1 s)$ and $a = 1$ otherwise.

Recursively, we can determine the sequence of n-stage value functions $\{V_1(s), V_2(s), \cdots V_n(s)\}$, and the limit of this sequence when n goes to infinity. Lippman in [10] shows that $V(s) := \lim_{n\to\infty} V_n(s)$ exists and it is the solution of the infinite horizon discounted cost problem. Besides, $V(s)$ is the unique solution to the dynamic programming equation (4).

IV. Structure of the optimal policy

The Policy Iteration algorithm [4] can be implemented to get numerically the optimal policy. In this study, we have implemented the algorithm by developing a C program. We define the two following examples with these parameters: $W = 10$, $\lambda_1 = \lambda_2 = 5$, $\mu_1 = \mu_2 = 1$, $k = 5$; The structure of the optimal policies are plotted for two cases: i) for $\beta = 0.1$ (Fig. 3) and ii) for $\beta = 0.5$ (Fig. 4). In both cases, the optimal policy has the well-known structure of the switching curve,

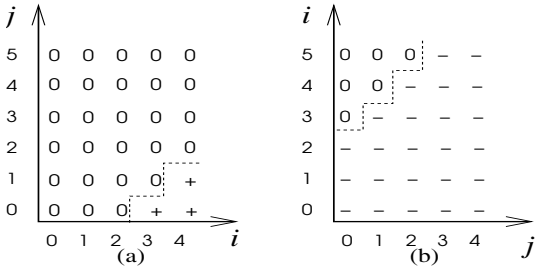

Fig. 3. Optimal Policy for $W = 10, k = 5, \lambda_1 = \lambda_2 = 5, \mu_1 = \mu_2 = 1, \beta = 0.1$ (Fig (a): $e = D_1$, Fig. (b): $e = D_2$).

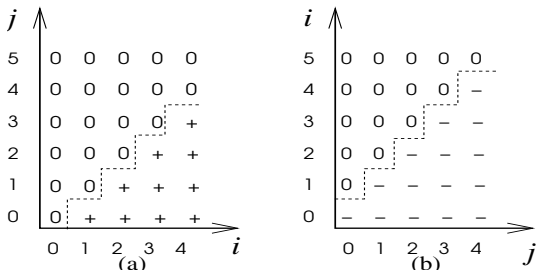

Fig. 4. Optimal Policy for $W = 10, k = 5, \lambda_1 = \lambda_2 = 5, \mu_1 = \mu_2 = 1, \beta = 0.5$ (Fig. (a): $e = D_1$, Fig. (b): $e = D_2$).

separating the state space into two domains. One domain for which the optimal action is $a = +1$ (depicted by '+') and the other one for which the optimal decision is $a = 0$. Figs. 3(a) and 4(a) show the optimal policy after a Class 1 call terminates. Similarly in Figs. 3(b) and 4(b), when a Class 2 call departures, the switching curve separates the states whose optimal action is $a = -1$ (depicted by '-') from states where the optimal action is $a = 0$. Furthermore, the switching curve is non-decreasing. A complete proof of the optimal policy monotonicity is in [11], which is available upon request.

The comparison between Fig. 3 and Fig. 4 illustrates the sensitivity of the switching curve to the factor β. When β is close to 1, the priority is given to Class 2 calls. Therefore, the system is more likely to reserve a larger number of wavelengths to this class; either by transferring one wavelength from Class 1 (at a Class 1 call termination), or by maintaining the number of wavelengths to Class 2 calls (at a Class 2 call termination).

Now we add Class 3 users to the problem and determine numerically the structure of the optimal policy. The problem formulation can be extended to the three classes problem by introducing ℓ, the number of available wavelengths for Class 3 calls. The global system state space is given by: $S = \{s = (i, j, \ell, k) | 0 \le i, \ell \le W - k, 0 \le j \le k, 0 \le k \le W\}$; and the parameter $\ell = W - k - n_3$ corresponds to the number of available wavelengths for Class 3 users. The reward function $R_g(s_n)$ with 3-class problem is a generalization of $R(s_n)$ defined in (1). Thus, $R_g(s_n)$ is a weighted sum of n_1, n_2 and n_3:

$$R_g(s_n) = n_1 + \beta n_2 + \delta n_3 \qquad (5)$$

yields the generalized cost function $C_g(s_n)$ as:

$$C_g(s_n) = B_g . s_n \qquad (6)$$

where $B_g = (1, \beta, \delta, (1 + \delta - \beta))^T$ with $0 \le \beta, \delta \le 1$ as the respective weights assigned to Classes 2 and 3, respectively.

The n-stage finite-horizon value function for three classes can be derived by using $C_g(s_n)$. Refer to [11] for a complete description of this problem.

In this paper, we only show the structure of the optimal policy, that is calculated with Policy iteration algorithm. Fig. 5 depicts the optimal policy for 3 classes of traffic for $W = 10, \ k = 4, \ \lambda_1 = \lambda_2 = \lambda_3 = 5, \ \mu_1 = \mu_2 = \mu_3 = 1, \ \beta =$

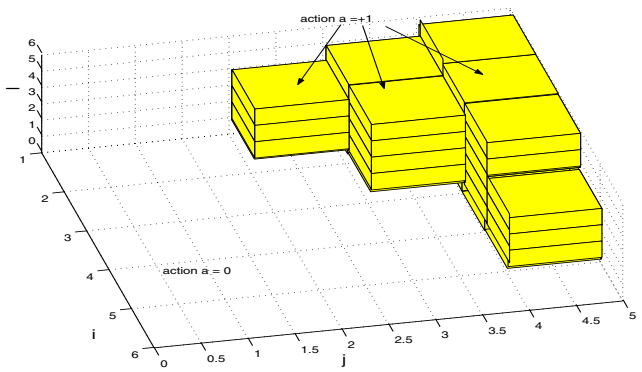

Fig. 5. Optimal Policy for 3 classes of traffic for $W = 10,\ k = 4,\ \lambda_1 = \lambda_2 = \lambda_3 = 5,\ \mu_1 = \mu_2 = \mu_3 = 1,\ \beta = 0.1,\ \delta = 0.1$.

$0.1,\ \delta = 0.1$, at a Class 1 call termination time. In this figure, each cube represents action $a = 1$. It can be noticed that the policy is a monotone 3D switching curve, dividing the state space into two subsets. The structure of the policy reflects the fact that the three classes of calls are competing for the available wavelengths. In Fig. 6, we set $\beta = 0.5$ and calculate the optimal policy for the same parameters as in Fig. 5. Comparison of Figs. 5 and 6 shows that by increasing β, the decision maker gives more resources to Class 2 users (i.e., more cubes).

V. Performance Comparison

In this section, we compare the performance of our proposed policy (DP), with Complete Sharing (CS) and Complete Partitioning (CP) policies. The simulation results are carried out only for traffic of Classes 1 and 2 with different weights. The performance metric used in the numerical comparison deals with: $n_1 + \beta n_2$.

In order to use CP policy, one can divide the total number of wavelengths, W, into two parts; each part corresponds to each class of traffic. Let m be the number of wavelengths allocated to Class 1 and consequently, $W - m$ wavelengths to Class 2. Note that m is constant. Using Erlang's B formula, we can compute p_k^i as the probability of having k users of Class i in the system. Then p_m^1 is the probability that m wavelengths are busy and being currently used by Class 1 users in the system. Using p_m^1 and p_{W-m}^2, we can derive the expected number of calls of each class as:

$$N_1(m) = \Big(\frac{\lambda_1}{\mu_1}\Big)(1 - p_m^1) \quad \text{and} \quad N_2(m) = \Big(\frac{\lambda_2}{\mu_2}\Big)(1 - p_{W-m}^2)$$

Let define m^* as:

$$m^* := \arg \max_{m \in \{1, \cdots W-1\}} N_1(m) + \beta N_2(m) \tag{7}$$

We simulate the system by deploying the optimal policy implemented in Section IV to calculate the performance metric of DP policy. For CP, the simulation result is carried out for two independent M/M/m^*/m^* and M/M/$W - m^*$/$W - m^*$ queues associated respectively with Class 1 and Class 2, where m^* is defined in (7). Finally, we simulate the system without any allocation policy in order to find the performance of CS policy.

Figs. 7 and 8 depict the average-time reward function ($n_1 + \beta n_2$) versus the load. In both examples, the parameters are set as follows: $W = 10$, $\lambda_1 = \lambda_2$ varying from 3 to 20, $\mu_1 = \mu_2 = 1$. The difference between the two examples lies in the value of β which is equal to 0.1 (respectively, 0.5) in Fig. 7 (respectively, Fig. 8). In Fig. 7, it can be observed that for low load (up to 6 Erlang), all the policies have a similar performance. As the load increases, DP policy shows much better performance, in particular when compared to CS policy.

Fig. 8 illustrates the case where β is equal to 0.5, thus closer to 1 compare with that in Fig. 7. One can notice that CP and DP policies yield to similar performance, which is better than CS policy performance. When β is close to 1, the priority assigned to Class 2 calls is close to the one assigned to Class 1 calls. Therefore, calls from both classes are equally competing for the access to the wavelengths. And CP, CS and DP policies have similar behavior in terms of resource allocation.

In order to evaluate the relative performance improvement of DP policy when compared to CS policy, we calculate the relative performance ratio, $(DP_p - CS_p)/CS_p$, where DP_p and CS_p represent the performance of DP and CS policies,

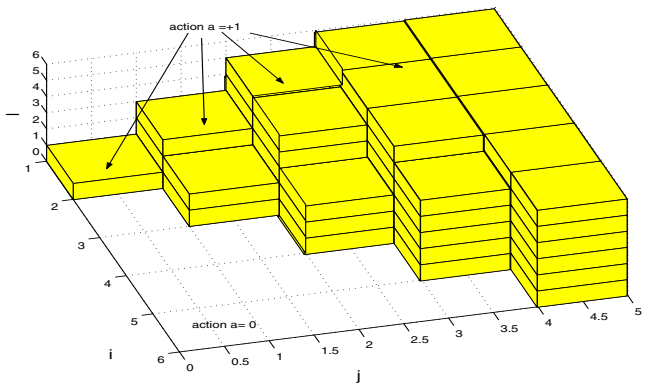

Fig. 6. Optimal Policy for 3 classes of traffic for $W = 10,\ k = 4,\ \lambda_1 = \lambda_2 = \lambda_3 = 5,\ \mu_1 = \mu_2 = \mu_3 = 1,\ \beta = 0.5,\ \delta = 0.1$.

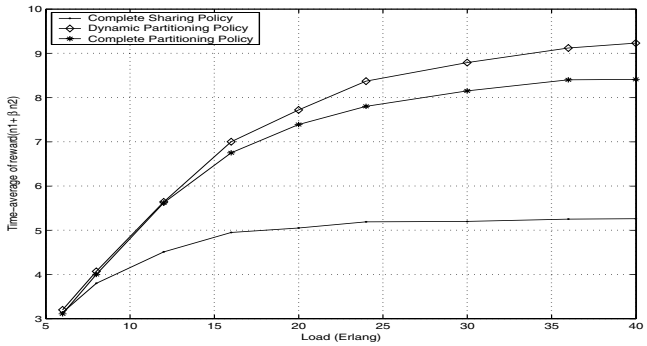

Fig. 7. Performance Comparison of DP, CP and CS policies for $W = 10, \lambda_1 = \lambda_2, \mu_1 = \mu_2 = 1, \beta = 0.1$.

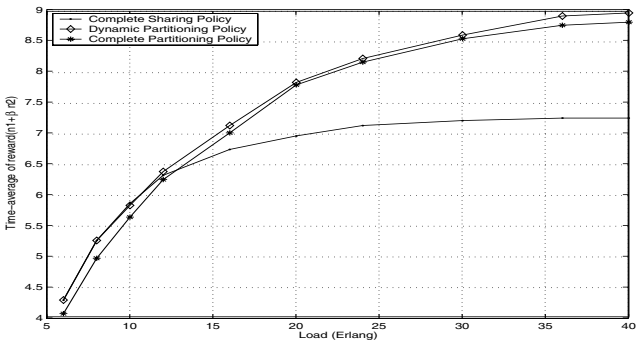

Fig. 8. Performance Comparison of DP, CP and CS policies for $W = 10, \lambda_1 = \lambda_2, \mu_1 = \mu_2 = 1, \beta = 0.5$.

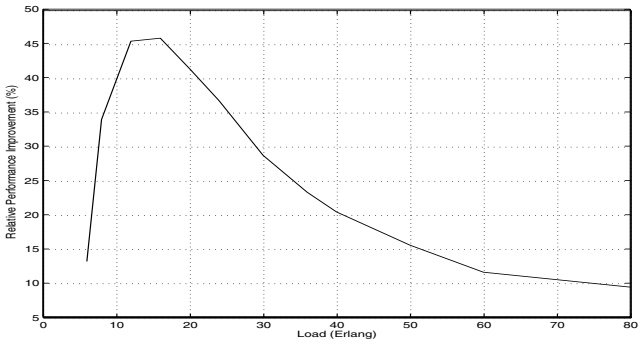

Fig. 10. Relative performance improvement of DP compared with CS when performance metric is the weighted blocking probabilities.

respectively. This quantity is plotted versus the load (Fig. 9). For $\beta = 0.1$, DP policy has significantly better performance, up to 75% for heavy load, whereas for $\beta = 0.5$, the best improvement is only equal to 25%. Another important performance metric is the weighted blocking probabilities. Using DP and CS policies, we simulate the system and determine this quantity. The relative performance improvement is plotted versus the load in Fig. 10. We can see that DP policy have higher performance, up to 45% for intermediate load (around 15 Erlang) which is a fact observed in networks and is in agreement with intuition.

Qualitatively, we can conclude that:
- when the two classes are differentiated with priorities, then DP policy is the most efficient policy.
- when the two classes are assigned the same weights, CS policy is a simpler policy with acceptable performance, in particular for network with low and intermediate load.

VI. Conclusion

We have described an approach to the problem of dynamic wavelength allocation in all-optical WDM networks. The problem has been formulated as an MDP and the optimal policy is obtained using the Policy Iteration method. The optimal policy which maximizes the reward function is a non-decreasing switching curve. The simulation results, carried out in a 2-hop network with two different classes with different weights, show that our policy provides the best performance in most cases. The investigation of the switching curve as a function of the parameters of the system would enable us to determine simple approximations of the optimal policy. Also, this method provides insight into the issue of resource utilization for more complex network topologies.

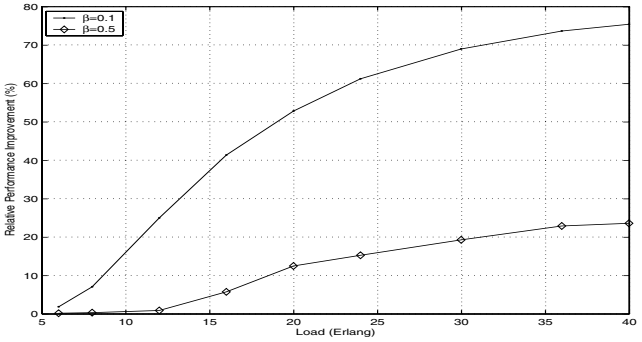

Fig. 9. Relative performance improvement of DP compared with CS when performance metric is the weighted sum of users.

References

[1] H. Zang, J. P. Jue, and B. Mukherjee, "A review of routing and wavelength assignment approaches for wavelength-routed optical WDM networks," *Optical Networks Magazine*, vol. 1, no. 1, Jan. 2000, pp. 47-60.

[2] H. Zang, J. P. Jue, and B. Mukherjee, "Dynamic lightpath establishment in wavelength-routed WDM networks," *IEEE Communications Magazine*, Sept. 2001, pp. 100-108.

[3] A. Jukan, H. R. van As, "Service-specific resource allocation in WDM networks with quality constraints," *IEEE Journal on Selected Area in Communications*, vol. 18, no. 10, Oct. 2000, pp. 2051-2061.

[4] M. L. Puterman, "Markov decision processes," *Wiley Inter-Science*, New York, 1994.

[5] I. Lambadaris, P. Narayan and I. Viniotis, "Optimal service allocatin amoung two heterogeneous traffic types with no queueing," in *Proc. of the 26th IEEE CDC*, Los Angeles, CA,Dec. 1987.

[6] J. Talim, Z. Liu and P. Nain, "Controlling robots of web search engines," in *Proc. of the ACM SIGMETRICS 2001 Performance 2001 Conference*, Cambridge, MA, June 2001.

[7] Y. Zhu, G. N. Rouskas and H. G. Perros, "A path decomposition approach for computing blocking probabilities in wavelength-routing networks," *IEEE/ACM Transactions on Networking*, vol. 8, No. 6, Dec. 2000.

[8] K. W. Ross, D. H. K. Tsang, "The stochastic knapsack problem," *IEEE Transactions on Communications*, vol. 37, no. 7, Jul. 1989, pp. 740-747.

[9] D. P. Bertsekas, "Dynamic progrmming deterministic and stochastic models," *Prentice-Hall*, Inc., Englewood cliffs, 1987.

[10] S. A. Lippman, "Semi-markov decision processes with unbounded rewards," *Management Science*, vol. 13, 1973, pp. 717-731.

[11] K. Mosharaf, I. Lambadaris, J. Talim, "On optimal resource allocation in all-optical WDM networks," Carleton University technical report, SCE-03-14, June 2003.