

Optimal Admission and Service Control in a Discrete time Service Facility Systems: MDP Approach

P. Maheswari¹ and C. Elango²

Research Department of Mathematical Sciences

Cardamom Planters' Association College

Bodinayakanur- 625 513, Tamil Nadu.

¹Corresponding Author. E-mail: maheswari23061993@gmail.com

²E-mail: chellaelango@gmail.com

Received 15 November 2017; accepted 10 December 2017

Abstract. In this article, we consider a discrete time service facility system under MDP structure. Decisions are taken at discrete time epochs to control both admission and service processes in the service facility system. Here the queue before the server is divided into *eligible queue* (with finite capacity N) and an unlimited *potential queue*. The number of arrivals and service completion with rate b follow general distributions with probability mass functions (p.m.f) $p(\cdot)$ and $q_b(\cdot)$ respectively. Control systems are used to (i) transfer customers from potential queue to eligible queue and (ii) change the service rate depending on the number of customers in the system. The system is formulated as a Markov Decision Process and an optimal control policy is obtained using Policy Iteration Method. Numerical example is provided to illustrate the problem with managerial insight.

Keywords: Markov decision processes, service control, admission control, discrete-time service facility system, reward analysis.

AMS Mathematics Subject Classification (2010): 60K25, 60K20

1. Introduction

Markov decision Process is a versatile and powerful tool for analyzing probabilistic sequential decision models with finite/ infinite planning horizon. MDP is a fusion of two concepts Markov Process and Dynamic programming.

Last three decades, many researchers in the field of operations and resource management contributed many results (Berman and Sapna [3], Berman and Kim [2], Arivarigan [1], Elango [5], Krishnamoorthy [8]). In most of the studies mentioned above, the system is considered as a Markov process with finite or infinite state space. The expressions for transition probability functions and the infinitesimal generator matrix of the Markov process are derived. The steady state probability distribution of the states has been found. Then by computing proper system performance measures and imposing respective cost structure, the cost analysis is done to get the optimal parameters of the system.

So far in the literature on discrete/continuous MDP models only admission control or service rate control problems are studied. We believe that an integrated approach like Markov Decision Process model is most appropriate to study service facility system (Queues- Inventory) and Machine maintenance systems (Machine-Spares). Berman and Sapna [4] studied one such service facility–inventory system under MDP structure using LPP method to control the service rates. Hild Mohamed et al. [6] analyzed a Markov decision problem: Optimal control of servers in a service facility holding perishable inventory with impatient customers.

In this paper, we imposed the Markov Decision Process (MDP) frame on a simple service facility (Queue system) to implement sequential decision making on both *admission* and *service rate*. This kind of decision problems arrive in feed back control of engineering systems, portfolio management and supply chain management etc. The standard mathematical formulation of this problem involves MDPs. Thus the states of the system is modeled as Markov chain, whose transitions probabilities depends on the appropriate action choices, by considering the state action dependent reward incurred at each stage.

Recently Kim [7] considered the admission control and the inventory management problem of a make-to-order (MTO) facility with a common component, which is purchased from a supplier under stochastic lead time with setup cost. Arriving demands of MTO type (customized types) are satisfied by using common (single) component. Selvakumar et al. [11] considered a discrete time MDP in a service facility system in which inventory is maintained to complete the service. Decisions are taken at discrete time epochs to control both admission and inventory control in service facility systems. Control system is used to transfer customers from potential queue to eligible queue, but with single demand class.

In this paper, we try to control both the admission and service in a service facility system under periodic review (equally spaced time epochs). The queue before the server is divided into *eligible queue* and *potential queue*. Here, we use policy iteration method to optimize the expected total reward. In the last section a numerical example is provided to illustrate the model.

2. Model description

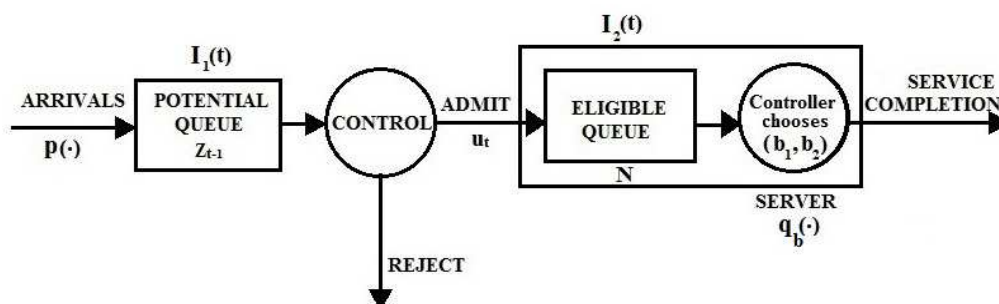


Figure 1: A service facility system with admission and service control

We formulate the model as follows: Decision epochs of the system correspond to the beginning of each period (Figure 2). The system is observed every $\eta > 0$ unit of time

Optimal Admission and Service Co Optimal Admission and Service Control in a Discrete time Service Facility Systems: MDP Approach

and the decision epochs are $\eta, 2\eta, \dots, L\eta$, $L < \infty$ (finite horizon). Admissions to the service facility is controlled, by observing the number of customers in the system (eligible queue + server). Service is controlled by selecting a service rate from the set of rates $B = \{b_1, b_2\}$. The service rates b_1 and b_2 denote the high and low efficiency servers respectively and they can be changed depending on the number of customers in the system. Assume that the maximum capacity of waiting space is N (finite). Arriving number of customers to service facility system follows a probability mass function $p(\cdot)$ and the arriving customers are placed in *potential queue*. Possible number of service completion follows a general probability mass function $q_b(\cdot)$ with rate b . The controller uses $q_b(\cdot)$ in period t and uses $q_{b'}(\cdot)$ in period $t+1$ means that a server change occurs. Maximum number of customers to be admitted at time epoch $t =$ maximum capacity of waiting space (N) - Number of customers in the eligible queue. Remaining customers are assumed to be rejected. All serviced customers depart the system at the end of period.

3. Main results

3.1. MDP formulation

We consider the problem as MDP having five components (tuples) $(T, S, A_s, p_t(\cdot | \cdot), r_t(\cdot))$.

Decision Epochs: $T = \{\eta, 2\eta, \dots, L\eta\}$, $L < \infty$ (finite).

State Space:

$$S = S_1 \times S_2 \times S_3 = \{0, 1, 2, \dots\} \times \{0, 1, 2, \dots, N\} \times \{b_1, b_2\},$$

$S = \{(i_1, i_2, i_3) : i_1 \text{ denotes the number of customers in the potential queue,}$

$i_2 \text{ denotes the number of customers in the system (eligible queue + server),}$

$i_3 \text{ denotes the current service rate}\}$

Actions:

$$A_{(i_1, i_2, i_3)} = \{ \{ \alpha, \beta \} \in A_s, \alpha = \{0, 1, 2, \dots, N - i_2\}, \beta = \{0, 1\} \}, \text{ (Policy is state dependent).}$$

Decision for Admission Control: $\alpha = \{0, 1, 2, \dots, N - i_2\}$, (Policy is state dependent).

Decision for Service Control:

$$\beta = \left\{ \begin{array}{l} 0 - \text{no server change} \\ 1 - \text{change of server (lower} \leftrightarrow \text{higher)} \end{array} \right\}$$

Transition probability:

$$p_t(s' | s, a) = \left\{ \begin{array}{ll} q_b(i_2 + a - i'_2) p(i'_1) & \text{if } a + i_2 > i'_2 > 0 \\ \sum_{n=i_2+a}^N q_b(n) p(i'_1) & \text{if } i'_2 = 0, a + i_2 > 0 \\ p(i_1) & \text{if } i'_2 = a + i_2 = 0 \\ 0 & \text{if } i'_2 > a + i_2 \geq 0. \end{array} \right.$$

$$i_1, i_2 \geq 0, i'_1, i'_2 \geq 0, s = (i_1, i_2, i_3), s' = (i'_1, i'_2, i'_3).$$

Rewards:

The *expected reward* as a function of state and action is

$$r_t(s, a) = r \times E\{\min(Y_t, i_2 + a)\} - w(i_2 + a) - c(b, b'), \quad a \in A = \bigcup_{s \in S} A_s, \quad s = (i_1, i_2, i_3).$$

The expected number of service completion in period t is given by

$$E\{\min(Y_t, i_2 + a)\} = \sum_{n=1}^{i_2+a-1} n \cdot q_b(n) + (i_2 + a) \sum_{n=i_2+a}^N q_b(n).$$

and the server changing cost is given by

$$c(b', b) = \begin{cases} K + d(b') & b \neq b' \\ d(b) & b = b' \end{cases}, \quad b = \{b_1, b_2\}, \quad b' = \{b'_1, b'_2\}.$$

The stationary reward structure consist of three components: the reward r for service completions and a waiting cost $w(i_2 + a)$ per period when there are $(i_2 + a)$ customers in the system (eligible queue) and a service rate cost $d(b)$ per period for using server $b(b_1 \text{ or } b_2)$ and a fixed cost K per period for changing service rate.

3.2. Analysis

Clearly $\{(I_1^{(t)}, I_2^{(t)}, I_3^{(t)}): t=0,1,\dots,L\}$ is a Markov chain with state space $S = \{0,1,2,\dots\} \times \{0,1,2,\dots,N\} \times \{b_1, b_2\}$. Let Z_t is the number of customers arrive during the period t . Customers arriving during the period $t - 1$ enter the potential queue. Let $I_1^{(t)}$ denote the number of customers in the potential queue, $I_2^{(t)}$ denote the number of customers in the system and $I_3^{(t)}$ current service rate, immediately prior to the decision epoch t . Let Y_t denote the number of “possible service completions” during period t .

At the decision epoch t , the controller admits $(N - I_2^{(t)})^+$ (number of waiting space in the system at time epoch t) = u_t of customers from the potential queue into the system, $t = 1, 2, \dots, L$.

$$(x)^+ = \begin{cases} x & \text{if } x > 0 \\ 0 & \text{if } x \leq 0. \end{cases}$$

Time	Potential Queue	System
t	$Z_{t-1} = I_1^{(t)}$	$I_2^{(t)}$
$t +$	0	$I_2^{(t)} + u_t$
$t + 1$	$Z_t = I_1^{(t+1)}$	$I_2^{(t+1)} = \begin{cases} I_2^{(t)} + I_1^{(t)} - Y_t & \text{if } I_1^{(t)} < N - I_2^{(t)} \\ N - Y_t & \text{if } I_1^{(t)} \geq N - I_2^{(t)} \end{cases}$

Table 1: State of the service facility system

Hence $t +$ denotes a point in time immediately after the control has been implemented but prior to any service completion.

Optimal Admission and Service Co Optimal Admission and Service Control in a Discrete time Service Facility Systems: MDP Approach

We can admit only $u_t = N - I_2^{(0)}$ customers, so that $0 \leq u_t \leq N$, $t = 1, 2, \dots, L$.

The random variable Y_t takes non-negative integer values and follows a time invariant probability mass function $q_b(n) = \Pr\{Y_t = n\}$, $t = 1, 2, \dots, L$ and Z_t assumes non-negative values which follows a time invariant probability mass function $p(n) = \Pr\{Z_t = n\}$, $t = 0, 1, 2, \dots, L$.

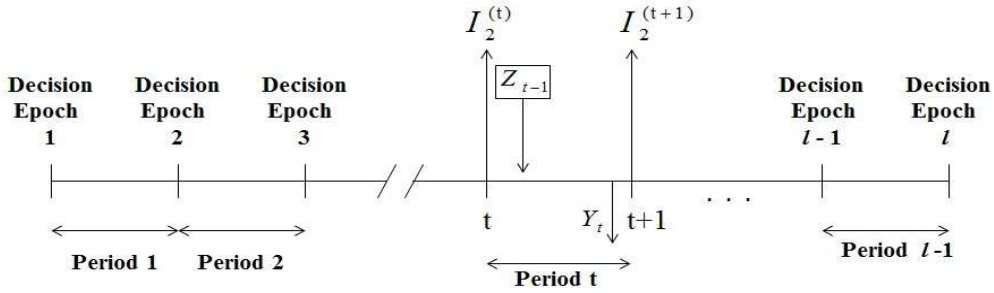


Figure 2: Functional relation diagram

The one step reward are given by, $r_t(s, a)$, where $s = (i_1, i_2, i_3)$ denote the state of the system at decision epoch t (beginning of t^{th} period). Assume the stationary policy R and hence the transition probability

$$p_t(s' | s, a) = \Pr\{(I_1^{(t+1)}, I_2^{(t+1)}, I_3^{(t+1)}) = s' | (I_1^{(t)}, I_2^{(t)}, I_3^{(t)}) = (s, a)\} \text{ where}$$

$s' = (i'_1, i'_2, i'_3)$, $s = (i_1, i_2, i_3)$, regardless the past history of the system up to time epoch t .

Then $\{(I_1^{(t)}, I_2^{(t)}, I_3^{(t)}) : t = 0, 1, \dots, L\}$ is a Markov chain with discrete state space

$S = S_1 \times S_2 \times S_3$. The t -step transition probabilities of the Markov chain under policy R is given by

$$p^{(t)}(s' | s)(R) = \Pr\{(I_1^{(t)}, I_2^{(t)}, I_3^{(t)}) = s' | (I_1^{(0)}, I_2^{(0)}, I_3^{(0)}) = s\},$$

Let $V_t(s, R)$, denote the total expected reward over the first t decision epochs with initial state (i_1, i_2, i_3) when policy R is adopted.

Then

$$V_t(s, R) = \sum_{k=0}^{t-1} \sum_{s' \in S} p^{(k)}(s' | s)(R) r_{s'}(R, s'), \quad s' = (i'_1, i'_2, i'_3), \quad s = (i_1, i_2, i_3)$$

where

$$r_s(R) = r \cdot \phi_r - w \times \bar{L} - c \cdot \bar{\alpha}$$

r - reward for service completion; w - waiting cost of customer/period; $c(\cdot, \cdot)$ - service cost for server per period; ϕ_r - number of customers served per period; \bar{L} - number of customers in the eligible queue + 1 in service counter; $\bar{\alpha}$ - denote the current service rate.

3.3. Reward analysis

The average reward function $g_s(R)$ is given by

$g_s(R) = \lim_{t \rightarrow \infty} \frac{1}{t} V_t(s, R)$, $(i_1, i_2, i_3) \in S$. The elements of the above average reward function is due to the Theorem (Puterman [9] and Tijms [10]).

Theorem 3.3.1.

For all $s' = (i'_1, i'_2, i'_3)$, $s = (i_1, i_2, i_3) \in S$, $\lim_{t \rightarrow \infty} \frac{1}{t} \sum_{k=1}^t p_t^{(k)}(s' | s)(R)$ always exists and.

$$\lim_{t \rightarrow \infty} \frac{1}{t} \sum_{k=1}^t p^{(k)}(s' | s) = \begin{cases} \frac{1}{\mu_{s'}} & \text{if state } s' \text{ is recurrent} \\ 0 & \text{if state } s' \text{ is transient,} \end{cases}$$

where $\mu_{s'}$ denote the mean recurrent time from state (i'_1, i'_2, i'_3) to itself.

$$\text{Also } \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{k=1}^t p^{(k)}(s' | s) = f_{(s)}^{(s')} \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{k=1}^t p_t^{(k)}(s').$$

Since the Markov Chain $\{(I_1^{(t)}, I_2^{(t)}, I_3^{(t)}) : t = 0, 1, 2, \dots, L\}$ is a unichain, irreducible, all its states are Ergodic and have a unique equilibrium distribution.

$$\text{Thus, } \pi_{(s')} (R) = \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{k=1}^t p^{(k)}(s' | s)(R), \text{ exist and is independent of initial state,}$$

such that $\pi P = \pi$ and $\sum_{s \in S} \pi_{(s)} = 1$.

3.4. Optimal policy

A stationary policy R^* is said to be an average reward optimal policy if $g_{i_1, i_2, i_3}(R^*) \leq g_{i_1, i_2, i_3}(R)$ for each stationary policy R uniformly with the initial state (i_1, i_2, i_3) .

The relative value associated with a given policy R provides a tool for constructing anew policy R^* whose average reward is more than that of the current policy R .

The objective is to improve the given policy R whose average reward is $g(R)$ and relative value $v_{(i_1, i_2, i_3)}(R)$, $(i_1, i_2, i_3) \in S$.

By constructing a new policy R such that for each $s = (i_1, i_2, i_3) \in S$,

$$r_s(R^*) - g(R) + \sum_{s' \in S} p_{(s, s')} (R^*) v_{s'} \leq v_s \quad (1)$$

, where $s' = (i'_1, i'_2, i'_3)$, we obtain an improved policy R^* with $g(R^*) \leq g(R)$. We have to find the optimal policy R^* satisfying (1) which maximizes the reward functions

$$r_i(a) - g(R) + \sum_{s' \in S} p_t(s' | s, a) v_{s'}(R) \text{ over all actions } a \in A(s).$$

3.5. Algorithm

Step 0: (Initialization)

Choose a stationary policy R for the periodic review based admission and service control in service facility system with inventory.

Step 1: (Value determination step)

For the current policy R , compute the unique solution $(g(R), v_s(R))$ to the following linear equations

$$v_s = r_s(R) - g(R) + \sum_{s' \in S} p_t(s'|s)(R)v_{s'}(R), \quad s = (i_1, i_2, i_3) \in S,$$

$v_i = 0$, where i is an arbitrarily chosen state in S .

Step 2: (Policy improvement)

For each state $s = (i_1, i_2, i_3) \in S$ determine the actions yielding, optimal reward, that is

$$a^* \in \arg \max_{a \in A_s} \left\{ r_s(a) - g(R) + \sum_{s' \in S} p_t(s'|s, a)v_{s'}(R) \right\}.$$

The new stationary policy R^* is obtained by choosing $R^* = a_s^*$.

Step 3: (Convergence test)

If the new policy $R^* = R$ (the old one), then the searching process stops with policy R . Otherwise go to **Step 1** with R replaced by new R^* .

8. Numerical example

Consider a MDP formulation of a service facility system under periodic review which controls the transfer of customers from potential queue to eligible queue and change the service rate. Decisions at equidistant time epochs are taken to admit the eligible number of customers and select the service rate depending on the number of customers in the system.

For the system, we assume, $N = 5$. The state space become

$$S = \{(5, b_1), (5, b_2), (4, b_1), (4, b_2), (3, b_1), (3, b_2), (2, b_1), (2, b_2), (1, b_1), (1, b_2), (0, b_1), (0, b_2)\}.$$

Action set at $(i_1, i_2, i_3) \in S$ is

$$A_{(i_1, i_2, i_3)} = \{ \{ \alpha, \beta \} \in A_s, \alpha = \{0, 1, 2, \dots, N - i_2\}, \beta = \{0, 1\} \},$$

Assume the reward $r=1$ for service completion per customer, waiting cost $w=0.35$ per customer/ period, service rate cost: $d(b_1)=1$ for using server b_1 , and $d(b_2)=0.7$ for using server b_1 per period. A fixed cost $K=0.4$ per period is assumed for changing service rate (b_1 to b_2 or b_2 to b_1).

Computational procedure

For any given policy R , the policy improvement quantity is given by

$$T_s(a, R) = r_s(a) - g(R) + \sum_{s' \in S} p_t(s' | s)(a) v_{s'}(a) \quad \text{where } T_s(a, R) = v_s(R) \text{ for } a = R_s.$$

u	r(u)
0	1(0) = 0
1	1(0 x 0.20 + 1 x 0.80) = 0.80
2	1(0 x 0.02 + 1 x 0.43 + 2 x 0.55) = 1.53
3	1(0 x 0.08 + 1 x 0.26 + 2 x 0.34 + 3 x 0.32) = 1.90
4	1(0 x 0.07 + 1 x 0.19 + 2 x 0.22 + 3 x 0.28 + 4 x 0.24) = 2.53
5	1(0 x 0.07 + 1 x 0.19 + 2 x 0.22 + 3 x 0.28 + 4 x 0.24) = 2.53

Table 2: Expected reward for service completion using service rate b_1

u	r(u)
0	1(0) = 0
1	1(0 x 0.40 + 1 x 0.60) = 0.60
2	1(0 x 0.25 + 1 x 0.40 + 2 x 0.35) = 1.1
3	1(0 x 0.10 + 1 x 0.31 + 2 x 0.30 + 3 x 0.29) = 1.78
4	1(0 x 0.09 + 1 x 0.29 + 2 x 0.27 + 3 x 0.21 + 4 x 0.14) = 2.02
5	1(0 x 0.09 + 1 x 0.29 + 2 x 0.27 + 3 x 0.21 + 4 x 0.14) = 2.02

Table 3: Expected reward for service completion using service rate b_2

$s \setminus s'$	(5,b ₁)	(5,b ₂)	(4,b ₁)	(4,b ₂)	(3,b ₁)	(3,b ₂)	(2,b ₁)	(2,b ₂)	(1,b ₁)	(1,b ₂)	(0,b ₁)	(0,b ₂)
(5,b ₁)	0.04	0	0.06	0	0.14	0	0.21	0	0.30	0	0.25	0
(5,b ₂)	0	0.18	0	0.30	0	0.25	0	0.15	0	0.07	0	0.05
(4,b ₁)	0	0	0.07	0	0.10	0	0.20	0	0.35	0	0.28	0
(4,b ₂)	0	0	0	0.17	0	0.29	0	0.25	0	0.19	0	0.10
(3,b ₁)	0	0	0	0	0.14	0	0.22	0	0.40	0	0.24	0
(3,b ₂)	0	0	0	0	0	0.18	0	0.35	0	0.28	0	0.19
(2,b ₁)	0	0	0	0	0	0	0.18	0	0.45	0	0.37	0
(2,b ₂)	0	0	0	0	0	0	0	0.20	0	0.44	0	0.36
(1,b ₁)	0	0	0	0	0	0	0	0	0.38	0	0.62	0
(1,b ₂)	0	0	0	0	0	0	0	0	0	0.4	0	0.6
(0,b ₁)	0	0	0	0	0	0	0	0	0	0	1.0	0
(0,b ₂)	0	0	0	0	0	0	0	0	0	0	0	1.0

Table 4: Pre-specified transition probabilities of the system

Optimal Admission and Service Co Optimal Admission and Service Control in a Discrete time Service Facility Systems: MDP Approach

Iteration 1:

Policy iteration algorithm is initialized with

$$R^{(1)} = (\{0,0\}, \{0,0\}, \{0,0\}, \{0,0\}, \{0,0\}, \{0,0\}, \{0,0\}, \{0,0\}, \{0,0\}, \{0,0\}, \{4,0\}, \{3,0\}),$$

which prescribes the transfer of 4 customers from potential queue to the system(eligible queue + 1 in server) in state $(0, b_1)$ and 3 customers at state $(0, b_2)$ respectively. Solving the system of linear equations connecting the average reward $g(R)^{(1)}$ by assuming $v_{0i}, i=1,2$ we get

$$\begin{aligned} v_{(5,b_1)}(R^{(1)}) &= -.9826193917, v_{(5,b_2)}(R^{(1)}) = -1.270295063, v_{(4,b_1)}(R^{(1)}) = -.6050411108, \\ v_{(4,b_2)}(R^{(1)}) &= -.7950209374, v_{(3,b_1)}(R^{(1)}) = -.8381333138, v_{(3,b_2)}(R^{(1)}) = -.6370426829, \\ v_{(2,b_1)}(R^{(1)}) &= -.7572777341, v_{(2,b_2)}(R^{(1)}) = -.8525000000, v_{(1,b_1)}(R^{(1)}) = -.935483871, \\ v_{(1,b_2)}(R^{(1)}) &= -.8000000000, v_{(0,b_1)}(R^{(1)}) = 0, v_{(0,b_2)}(R^{(1)}) = 0, g(R^{(1)}) = .030000000. \end{aligned}$$

$s \setminus \alpha$	{5,0}	{5,1}	{4,0}	{4,1}	{3,0}	{3,1}	{2,0}	{2,1}	{1,0}	{1,1}	{0,0}	{0,1}
(5,b ₁)	x	x	x	x	x	x	x	x	x	x	-.9826193918	-.83
(5,b ₂)	x	x	x	x	x	x	x	x	x	x	-1.270295063	-.72
(4,b ₁)	x	x	x	x	x	x	x	x	-.32	-.83	-.6050411108	-.48
(4,b ₂)	x	x	x	x	x	x	x	x	-.43	-.72	-.7950209374	-.37
(3,b ₁)	x	x	x	x	x	x	-.32	-.83	.03	-.48	-.8381333138	-.37
(3,b ₂)	x	x	x	x	x	x	-.43	-.72	-.08	-.37	-.6370426829	-.55
(2,b ₁)	x	x	x	x	-.32	-.83	.03	-.48	-.15	-.37	-.7572777341	-.70
(2,b ₂)	x	x	x	x	-.43	-.72	-.08	-.37	.03	-.55	-.8525000000	-.57
(1,b ₁)	x	x	-.32	-.83	.03	-.48	-.15	-.37	-.17	-.70	-.9354838710	-.085
(1,b ₂)	x	x	-.43	-.72	-.08	-.37	.03	-.55	-.30	-.57	-.8000000000	-.95
(0,b ₁)	-.62	-1.01	.03	-.48	-.15	-.37	-.17	-.70	-.55	-.85	-1.0300000000	-1.1
(0,b ₂)	-.43	-.72	-.08	-.37	.03	-.55	-.30	-.57	-.45	-.95	-.7300000000	-1.4

Table 5: Iteration 1: reward matrix for different decisions

The new policy will be

$$R^{(2)} = (\{0,1\}, \{0,1\}, \{1,0\}, \{0,1\}, \{1,0\}, \{1,0\}, \{2,0\}, \{1,0\}, \{3,0\}, \{2,0\}, \{4,0\}, \{3,0\}).$$

Since the new policy $R^{(2)}$ is different from the initial policy $R^{(1)}$, the searching process continues.

Iteration 2:

For the policy $R^{(2)}$, solving the system of linear equations connecting the average reward $g(R)^{(2)}$ by assuming $v_{0i}, i=1,2$ we get

P. Maheswari and C. Elango

$$v_{(5,b_1)}(R^{(2)}) = -.9193548387, v_{(5,b_2)}(R^{(2)}) = -1.148995148, v_{(4,b_1)}(R^{(2)}) = -.3763440860,$$

$$v_{(4,b_2)}(R^{(2)}) = -.5287981193, v_{(3,b_1)}(R^{(2)}) = 0, v_{(3,b_2)}(R^{(2)}) = -.1341463415, v_{(2,b_1)}(R^{(2)}) = 0,$$

$$v_{(2,b_2)}(R^{(2)}) = 0, v_{(1,b_1)}(R^{(2)}) = 0, v_{(1,b_2)}(R^{(2)}) = 0, v_{(0,b_1)}(R^{(2)}) = 0, v_{(0,b_2)}(R^{(2)}) = 0,$$

$$g(R^{(2)}) = .0300000000.$$

$s \setminus \{\alpha, \beta\}$	{5,0}	{5,1}	{4,0}	{4,1}	{3,0}	{3,1}	{2,0}	{2,1}	{1,0}	{1,1}	{0,0}	{0,1}
(5,b ₁)	x	x	x	x	x	x	x	x	x	x	-.4093548388	-.83
(5,b ₂)	x	x	x	x	x	x	x	x	x	x	-.8589951478	-.72
(4,b ₁)	x	x	x	x	x	x	x	x	-.32	-.83	-.0263440860	-.48
(4,b ₂)	x	x	x	x	x	x	x	x	-.43	-.72	-.2387981193	-.37
(3,b ₁)	x	x	x	x	x	x	-.32	-.83	.03	-.48	-.1800000000	-.37
(3,b ₂)	x	x	x	x	x	x	-.43	-.72	-.08	-.37	-.0241463414	-.55
(2,b ₁)	x	x	x	x	-.32	-.83	.03	-.48	-.15	-.37	-.2000000000	-.70
(2,b ₂)	x	x	x	x	-.43	-.72	-.08	-.37	.03	-.55	-.3300000000	-.57
(1,b ₁)	x	x	-.32	-.83	.03	-.48	-.15	-.37	-.17	-.70	-.5800000000	-.085
(1,b ₂)	x	x	-.43	-.72	-.08	-.37	.03	-.55	-.30	-.57	-.4800000000	-.95
(0,b ₁)	-.62	-1.01	.03	-.48	-.15	-.37	-.17	-.70	-.55	-.85	-1.0300000000	-1.1
(0,b ₂)	-.43	-.72	-.08	-.37	.03	-.55	-.30	-.57	-.45	-.95	-.7300000000	-1.4

Table 6: Iteration 2: reward matrix for different decisions

The new policy will be

$R^{(3)} = (\{0,0\}, \{0,1\}, \{0,0\}, \{0,0\}, \{1,0\}, \{0,0\}, \{2,0\}, \{1,0\}, \{3,0\}, \{2,0\}, \{4,0\}, \{3,0\})$. Since the new policy $R^{(3)}$ is different from the initial policy $R^{(2)}$, the searching process continues.

Iteration 3:

For the policy $R^{(3)}$, solving the system of linear equations connecting the average reward $g(R)^{(2)}$ by assuming $v_{0i}, i = 1, 2$ we get

$$v_{(5,b_1)}(R^{(3)}) = -.3645833333, v_{(5,b_2)}(R^{(3)}) = -.9631207758, v_{(4,b_1)}(R^{(3)}) = 0,$$

$$v_{(4,b_2)}(R^{(3)}) = -.1325301205, v_{(3,b_1)}(R^{(3)}) = 0, v_{(3,b_2)}(R^{(3)}) = 0, v_{(2,b_1)}(R^{(3)}) = 0,$$

$$v_{(2,b_2)}(R^{(3)}) = 0, v_{(1,b_1)}(R^{(3)}) = 0, v_{(1,b_2)}(R^{(3)}) = 0, v_{(0,b_1)}(R^{(3)}) = 0, v_{(0,b_2)}(R^{(3)}) = 0,$$

$$g(R^{(3)}) = .030000000000.$$

$s \setminus \{\alpha, \beta\}$	{5,0}	{5,1}	{4,0}	{4,1}	{3,0}	{3,1}	{2,0}	{2,1}	{1,0}	{1,1}	{0,0}	{0,1}
(5,b ₁)	x	x	x	x	x	x	x	x	x	x	-.3645833333	-.83

Optimal Admission and Service Co Optimal Admission and Service Control in a Discrete time Service Facility Systems: MDP Approach

(5,b ₂)	x	x	x	x	x	x	x	x	x	x	-0.7631207758	-0.72
(4,b ₁)	x	x	x	x	x	x	x	x	-0.32	-0.83	0	-0.48
(4,b ₂)	x	x	x	x	x	x	x	x	-0.43	-0.72	-0.1325301205	-0.37
(3,b ₁)	x	x	x	x	x	x	-0.32	-0.83	0.03	-0.48	-0.1800000000	-0.37
(3,b ₂)	x	x	x	x	x	x	-0.43	-0.72	-0.08	-0.37	0	-0.55
(2,b ₁)	x	x	x	x	-0.32	-0.83	0.03	-0.48	-0.15	-0.37	-0.2000000000	-0.70
(2,b ₂)	x	x	x	x	-0.43	-0.72	-0.08	-0.37	0.03	-0.55	-0.3300000000	-0.57
(1,b ₁)	x	x	-0.32	-0.83	0.03	-0.48	-0.15	-0.37	-0.17	-0.70	-0.5800000000	-0.085
(1,b ₂)	x	x	-0.43	-0.72	-0.08	-0.37	0.03	-0.55	-0.30	-0.57	-0.4800000000	-0.95
(0,b ₁)	-0.62	-1.01	0.03	-0.48	-0.15	-0.37	-0.17	-0.70	-0.55	-0.85	-0.1030000000	-1.1
(0,b ₂)	-0.43	-0.72	-0.08	-0.37	0.03	-0.55	-0.30	-0.57	-0.45	-0.95	-0.7300000000	-1.4

Table 7: Iteration 3: reward matrix for different decisions

Since the new policy

$R^{(4)} = (\{0,0\}, \{0,1\}, \{0,0\}, \{0,0\}, \{1,0\}, \{0,0\}, \{2,0\}, \{1,0\}, \{3,0\}, \{2,0\}, \{4,0\}, \{3,0\})$ is identical with the policy, the searching process stops here. After three iterations we obtained the following optimal policy:

$$R^* = (\{0,0\}, \{0,1\}, \{0,0\}, \{0,0\}, \{1,0\}, \{0,0\}, \{2,0\}, \{1,0\}, \{3,0\}, \{2,0\}, \{4,0\}, \{3,0\}).$$

The optimal solution schedule is

$$\begin{array}{cccccccccccc}
 (5,b_1) & (5,b_2) & (4,b_1) & (4,b_2) & (3,b_1) & (3,b_2) & (2,b_1) & (2,b_2) & (1,b_1) & (1,b_2) & (0,b_1) & (0,b_2) \\
 \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\
 \{0,0\} & \{0,1\} & \{0,0\} & \{0,0\} & \{1,0\} & \{0,0\} & \{2,0\} & \{1,0\} & \{3,0\} & \{2,0\} & \{4,0\} & \{3,0\}
 \end{array}$$

9. Conclusion and future research

In this article we presented an application of Markov Decision Process (MDP) for admission and service control using classical approach namely 'policy iteration'. The optimum admission of customers and service rates to be employed is found so that expected reward is maximized. We are currently studying Markov Decision Process in discrete time with admission and service control. In future we would like to extend the model to control both service and replenishment order simultaneously in a service facility with inventory management..

Acknowledgement. P.Maheswari's research is supported by the University Grants Commission, Govt. of India, under NFOBC Scheme (F./2015-16/NFO-2015-17-OBC-TAM-46773/(SA-III/Website)).

REFERENCES

1. G.Arivarignan and B.Sivakumar, Inventory system with renewal demands at service facilities in S.K.Srinivasan and A.Vijayakumar (Eds.), *Stochastic point processes*, Narosa Publishing House, New Delhi, (2003) 108-123.
2. O.Berman and E.Kim, Dynamic inventory strategies for profit maximization in a service facility with stochastic service, *Math. Methods. Operation Research*, 60(3) (2004) 497-521.

P. Maheswari and C. Elango

3. O.Berman and K.P.Sapna, Inventory management at service facility for systems with arbitrarily distributed service times, *Commun. Stat. Stoch. Models*, 16(3) (2000) 343-360.
4. O.Berman and K.P.Sapna, Optimal control of service for facilities holding inventory, *Comput. Oper. Res.*, 28 (5) (2001) 429-441.
5. C.Elango and R.Satheesh Kumar, Markov decision processes for service facility systems with perishable inventory, *International Journal of Computer Applications*, 9(4) (2010).
6. H.M.Al Hamadi, N.Sangeetha and B.Sivakumar, Optimal control of service parameter for a perishable inventory system maintained at service facility with impatient customers, *Ann. Oper. Res.* DOI: 10.1007/s10479-014-1627-1.
7. E.Kim and T.Park, Admission and inventory control of a single-component make-to-order production system with replenishment setup cost and lead time, *European Journal of Operational Research*. doi:10.1016/j.ejor.2016.04.021.
8. A.Krishnamoorthy, V.C.Narayanan, T.G.Deepak and P.Vineetha, Control policies for inventory with service time, *Stochastic Analysis and Applications*, 24(4) (2006) 889-899.
9. M.L.Puterman, Markov Decision Processes: Discrete Stochastic Dynamic programming, *John Wiley and Sons Inc*, New York (1994).
1. H.C.Tijjims, A First Course in Stochastic Models, *John Wiley and Sons Ltd*, England (2003).
2. C.Selvakumar, P.Maheswari and C.Elango, Discrete MDP problem with admission and inventory control in service facility systems, *International Journal of Computational and Applied Mathematics*, 12(1) (2017).