Chapter 5

OPTIMAL DESIGN AND CONTROL ASPECTS OF CERTAIN SINGLE SERVER PRIORITY QUEUEING MODELS

CHAPTER 5

OPTIMAL DESIGN AND CONTROL ASPECTS OF CERTAIN SINGLE SERVER PRIORITY QUEUEING MODELS

5.1 Introduction

The study of optimality of priority queueing models yields results which enhance the usefulness of these queueing models in their applications to the real life situations. When priority queueing models are considered, one naturally thinks interms of minimising some cost function with regard to delay for different types of units. In this Chapter, the optimal design aspects of a priority queueing model involving multiple service facilities and the optimal control policy for another single server priority queueing model involving general service times are discussed. In Section 5.2, the optimality of design parameters involved in a priority queueing model involving multiple service facilities is discussed with the help of a cost function involving idle time costs, holding costs and overhead costs. These results are reported in Muthu and Sampathkumar (1994a). Subsequently, in Section 5.3, an optimal turn-on policy is obtained in the case of a single server priority queueing model involving general service times. A particular case is subsequently discussed. These types of modelling will be useful in the maintenance of a

fleet of vehicles which consists of two categories with the availability of a preferential repair facility for one of the two categories.

5.2 Optimal design aspects of a priority queueing model involving multiple service facilities

We now proceed to discuss the optimality of design parameters involved in a priority queueing model involving multiple service facilities. We consider here a finite source queueing model involving u_1 high-priority units and u_2 low-priority units. Initially, the high-priority units get service through the u_1 servers available in the first service facility on 'first-come-first-served' basis at the rate β_1, while the low-priority units get service through the u_2 servers available in the second service facility at the rate β_2 on 'first-come-first-served' basis. Then, both the high-priority and the low-priority units wait infront of the third service facility and get served on preemptive priority basis through the single server available in that facility at the rates α_1 and α_2 respectively. Once a unit gets served at the third service facility, it returns to either the first service facility or the second service facility according to its priority type.

Let $P_{rs}(t)$ be the probability that there are r high-priority units and s low-priority units in the third service

facility at time t and let $P_{rs,cd}(\Delta t)$ be the one step transition probability that the system moves from the state (r,s) to the state (c,d) in the infinitesimal time interval of length Δt. In order to develop the governing equations of the system at time t, the possible occurrences of events in the interval $(t, t+\Delta t)$ shall be postulated as follows:

$$P_{r+1s,rs}(\Delta t) = \alpha_1\Delta t+O(\Delta t),\ r=0,\ldots,u_1-1;\ s=0,\ldots,u_2$$

$$P_{rs+1,rs}(\Delta t) = 0,\ r=1,\ldots,u_1;\ s=0,\ldots,u_2-1$$

$$= \alpha_2\Delta t+O(\Delta t),\ r=0;\ s=0,\ldots,u_2-1$$

$$P_{r-1s,rs}(\Delta t) = (u_1-r+1)\beta_1\Delta t+O(\Delta t),\ r=1,\ldots,u_1;\ s=0,\ldots,u_2$$

$$P_{rs-1,rs}(\Delta t) = (u_2-s+1)\beta_2\Delta t+O(\Delta t),\ r=0,\ldots,u_1;\ s=1,\ldots,u_2$$

$$P_{rs,r+m,s+n}(\Delta t) = O(\Delta t),\ m\geq 0,\ n\geq 0,\ m+n>1$$

Using the above assumptions, the governing equations of the system under consideration shall be obtained as follows:

$$P_{oo}(t+\Delta t) = P_{oo}(t)[1-u_1\beta_1\Delta t-u_2\beta_2\Delta t]+P_{1o}(t)[\alpha_1\Delta t]$$
$$+P_{o1}(t)[\alpha_2\Delta t]+o(\Delta t)$$

$$P_{os}(t+\Delta t) = P_{os}(t)[1-\alpha_2\Delta t-u_1\beta_1\Delta t-(u_2-s)\beta_2\Delta t]+$$
$$+P_{os+1}(t)[\alpha_2\Delta t]+P_{1s}(t)[\alpha_1\Delta t]$$
$$+P_{os-1}(t)[u_2-s+1)\beta_2\Delta t]+o(\Delta t)$$

$$P_{ou_2}(t+\Delta t) = P_{ou_2}(t)[1-\alpha_2\Delta t-u_1\beta_1\Delta t]+P_{1u_2}(t)[\alpha_1\Delta t]$$
$$+P_{ou_2-1}(t)[\beta_2\Delta t]+O(\Delta t)$$

$$P_{ro}(t+\Delta t) = P_{ro}(t)[1-\alpha_1\Delta t-(u_1-r)\beta_1\Delta t-u_2\beta_2\Delta t] + P_{r-1o}(t)[(u_1-r+1)\beta_1\Delta t]+P_{r+1o}(t)[\alpha_1\Delta t] + o(\Delta t)$$

$$P_{rs}(t+\Delta t) = P_{rs}(t)[1-\alpha_1\Delta t-(u_1-r)\beta_1\Delta t-(u_2-s)\beta_2\Delta t] + P_{r+1s}(t)[\alpha_1\Delta t]+P_{r-1s}(t)[(u_1-r+1)\beta_1\Delta t] + P_{rs-1}(t)[(u_2-s+1)\beta_2\Delta t]+o(\Delta t)$$

$$P_{ru_2}(t+\Delta t) = P_{ru_2}(t)[1-\alpha_1\Delta t-(u_1-r)\beta_1\Delta t]+P_{r+1u_2}(t)[\alpha_1\Delta t] + P_{r-1u_2}(t)[(u_1-r+1)\beta_1\Delta t]+P_{ru_2-1}(t)[\beta_2\Delta t] + o(\Delta t)$$

$$P_{u_1o}(t+\Delta t) = P_{u_1o}(t)[1-\alpha_1\Delta t-u_2\beta_2\Delta t]+P_{u_1-1o}(t)[\beta_1\Delta t]+o(\Delta t)$$

$$P_{u_1s}(t+\Delta t) = P_{u_1s}(t)[1-\alpha_1\Delta t-(u_2-s)\beta_2\Delta t]+P_{u_1-1s}(t)[\beta_1\Delta t] + P_{u_1s-1}(t)[(u_2-s+1)\beta_2\Delta t]+o(\Delta t)$$

$$P_{u_1u_2}(t+\Delta t)= P_{u_1u_2}(t)[1-\alpha_1\Delta t]+P_{u_1-1u_2}(t)[\beta_1\Delta t] + P_{u_1u_2-1}(t)[\beta_2\Delta t]+o(\Delta t)$$

Transposing, dividing through by Δt and taking limit as $\Delta t \to 0$, the following difference-differential equations shall be obtained.

$$(u_1\beta_1+u_2\beta_2)P_{oo}(t) = \alpha_1 P_{1o}(t) + \alpha_2 P_{o1}(t) \quad (5.2.1)$$

$$[\alpha_2+u_1\beta_1+(u_2-s)\beta_2]P_{os}(t) = \alpha_2 P_{os+1}(t) + \alpha_1 P_{1s}(t) + (u_2-s+1)\beta_2 P_{os-1}(t) \quad (5.2.2)$$

$$(\alpha_2+u_1\beta_1) P_{ou_2}(t) = \alpha_1 P_{1u_2}(t)+\beta_2 P_{ou_2-1}(t) \quad (5.2.3)$$

$$[\alpha_1+(u_1-r)\beta_1+u_2\beta_2]P_{ro}(t) = (u_1-r+1)\beta_1 P_{r-1o}(t) + \alpha_1 P_{r+1o}(t) \quad (5.2.4)$$

$$[\alpha_1+(u_1-r)\beta_1+(u_2-s)\beta_2]P_{rs}(t) = \alpha_1 P_{r+1s}(t) + (u_1-r+1)\beta_1 P_{r-1s}(t) + (u_2-s+1)\beta_2 P_{rs-1}(t) \quad (5.2.5)$$

$$[\alpha_1+(u_1-r)\beta_1]P_{ru_2}(t) = \alpha_1 P_{r+1u_2}(t)+(u_1-r+1)\beta_1 P_{r-1u_2}(t) + \beta_2 P_{ru_2-1}(t) \quad (5.2.6)$$

$$(\alpha_1+u_2\beta_2)P_{u_1o}(t) = \beta_1 P_{u_1-1o}(t) \quad (5.2.7)$$

$$[\alpha_1+(u_2-s)\beta_2]P_{u_1s}(t) = \beta_1 P_{u_1-1s}(t)+(u_2-s+1)\beta_2 P_{u_1s-1}(t) \quad (5.2.8)$$

$$\alpha_1 P_{u_1u_2}(t) = \beta_1 P_{u_1-1u_2}(t)+\beta_2 P_{u_1u_2-1}(t) \quad (5.2.9)$$

These difference-differential equations cannot be directly solved easily and hence we shall define and use the following generating functions to obtain the desired results.

Let $H_r(y,t) = \sum_{s=0}^{u_2} P_{rs}(t)\, y^s$

$$H(x,y,t) = \sum_{r=0}^{u_1} H_r(y,t)x^r$$

Multiplying both sides of (5.2.1),(5.2.2) and (5.2.3) by y^0, y^s and y^{u_2} respectively and summing over s, we obtain

$$\partial H_o(y,t)/\partial t = [\alpha_2((1/y)-1)-u_1\beta_1+u_2\beta_2(y-1)]H_o(y,t)$$
$$+\alpha_1 H_1(y,t)+\beta_2 y(1-y)\partial H_o(y,t)/\partial y$$
$$+\alpha_2(1-(1/y))P_{oo}(t) \qquad (5.2.10)$$

Multiplying both sides of (5.2.4), (5.2.5) and (5.2.6) by y^0, y^s and y^{u_2} respectively and summing over s, we obtain

$$\partial H_r(y,t)/\partial t = [-\alpha_1-u_1\beta_1+r\beta_1-u_2\beta_2+u_2\beta_2 y]\ H_r(y,t)$$
$$+(u_1-r+1)\beta_1 H_{r-1}(y,t)+\alpha_1 H_{r+1}(y,t)$$
$$+\beta_2 y(1-y)\partial H_r(y,t)/\partial y \qquad (5.2.11)$$

Multiplying both sides of (5.2.7), (5.2.8) and (5.2.9) by y^0, y^s and y^{u_2} respectively and summing over s, we obtain

$$\partial H_{u_1}(y,t)/\partial t = [-\alpha_1-u_2\beta_2+(u_2+1)\beta_2 y-\beta_2 y]H_{u_1}(y,t)$$
$$+\beta_1 H_{u_1-1}(y,t)+(\beta_2 y-\beta_2 y^2)\partial H_{u_1}(y,t)/\partial y$$
$$+u_2\beta_2 P_{u_1u_2}(t)y^{u_2}-u_2\beta_2 P_{u_1u_2}(t)\ y^{u_2}$$
$$-(u_2+1)\beta_2 P_{u_1u_2-1}(t)y^{u_2}-(u_2+1)\beta_2 P_{u_1u_2}(t)y^{u_2+1}$$
$$+u_2\beta_2 P_{u_1u_2-1}(t)\ y^{u_2}+(u_2+1)\beta_2 P_{u_1u_2}(t)y^{u_2+1}$$
$$+\beta_2\ P_{u_1u_2-1}(t)\ y^{u_2} \qquad (5.2.12)$$

Then, multiplying both sides of (5.2.10), (5.2.11) and (5.2.12) by x^0, x^r and x^{u_1} and summing over r, we obtain the following probability generating function

$$\partial H(x,y,t)/\partial t = [\alpha_1((1/x)-1)+u_1\beta_1(x-1)+u_2\beta_2(y-1)]H(x,y,t)$$
$$+[\alpha_1(1-(1/x))+\alpha_2((1/y)-1)]H_o(y,t)$$
$$+[u_1\beta_1 x^{u_1}(x-1)+\beta_1 x^{u_1}(1-x)]\ H_{u_1-1}(y,t)$$
$$+\beta_1 x(1-x)\ \partial H(x,y,t)/\partial x$$
$$+\beta_2 y(1-y)\ \partial H(x,y,t)/\partial y$$
$$+\alpha_2(1-(1/y))\ P_{oo}(t) \qquad (5.2.13)$$

In the steady state, the joint generating function shall be obtained in the following form.

$$[\alpha_1((1/x)-1)+u_1\beta_1(x-1)+u_2\beta_2(y-1)]\ H(x,y)$$
$$= [\alpha_1((1/x)-1)+\alpha_2(1-(1/y))]\ H_o(y)$$
$$+(u_1-1)(1-x)\beta_1 x^{u_1}\ H_{N_1-1}(y)$$
$$+\beta_1 x(x-1)\partial H(x,y)/\partial x + \beta_2 y(y-1)\partial H(x,y)/\partial y$$
$$+\alpha_2((1/y)-1)\ P_{oo} \qquad (5.2.14)$$

Differenting (5.2.14) with respect to x and substituting x=y=1 and using equations (5.2.1),(5.2.2) and (5.2.5),the steady-state expected number of priority-1 units shall be obtained as

$$E(x_1) = \sum_{k=1}^{u_1} (k\ u_1!\ \beta_1^k\ P_o)/((u_1-k)!\ \alpha_1^k)$$

$$\text{where } P_o = [\sum_{k=1}^{u_1} (u_1!\ \beta_1^k)/((u_1-k)!\ \alpha_1^k)]^{-1} \qquad (5.2.15)$$

Similarly, differentiating (5.2.14) with respect to y and substituting x=y=1, the steady-state expected number of priority-2 units shall be obtained as

$$E(X_2) = (\alpha_1\alpha_2u_2\beta_2-\alpha_1\alpha_2^2P_o+\alpha_2(\alpha_1\alpha_2-\alpha_2\beta_1-\alpha_1\beta_2)/(\alpha_1\alpha_2\beta_2) \quad (5.2.16)$$

Using c_s to denote the marginal cost of the server per unit time, K to denote a fixed cost per unit time, c_k to denote waiting time cost per priority-k (k=1,2) unit per unit time and α to denote the common service time, the expected total cost per unit time shall be expressed as

$$E(C) = K + c_s\alpha + c_1 \sum_{j=1}^{u_1} (ju_1!\beta_1^jP_o)/(u_1-j)!\ \alpha^j)$$

$$+C_2(\alpha^2u_2\beta_2-\alpha^3P_o+\alpha(\alpha^2-\alpha\beta_1-\alpha\beta_2))/(\alpha^2\beta_2) \quad (5.2.17)$$

It shall be shown that E(C) is a strictly convex function and setting $\partial E(C)/\partial\beta_k$ (k=1,2) and $\partial E(C)/\partial\alpha$ to zero yields equations in β_1, β_2 and α. The roots of these equations found by an appropriate numerical method provide the required optimal values of β_k (k=1,2) and α, denoted by β_1^*, β_2^* and α^*. The values of α^* for known values of c_s, c_1,c_2 and K and varying values of β_1 and β_2 are provided in the table 5.2.1. The values of β_1^* and β_2^* shall be obtained in a similar manner for different values of the other two parameters.

Table 5.2.1

Optimal values of α

	$C_s=1,\ C_1=2,\ C_2=1$					
	$\beta_2=0.05$	$\beta_2=0.06$	$\beta_2=0.07$	$\beta_2=0.08$	$\beta_2=0.09$	$\beta_1=0.10$
$\beta_1=0.05$	0.0883	0.1062	0.1231	0.1406	0.1728	0.2061
$\beta_1=0.06$	0.1119	0.1328	0.1576	0.1762	0.2018	0.2316
$\beta_1=0.07$	0.1472	0.1608	0.1924	0.2117	0.2334	0.2582
$\beta_1=0.08$	0.2062	0.2234	0.2407	0.2653	0.2918	0.3145

5.3 Optimal control aspects of a single server priority queueing model involving general service times

We now proceed to discuss the optimal control aspects of a nonpreemptive single server priority queueing model involving general service times.

We consider here a single server queueing model with two priority classes. Units of the high-priority and low-priority classes arrive according to independent Poisson processes, the mean arrival rate for class-k being α_k^{-1} (k=1,2). Service times for priority-k units are independently and identically distributed with distribution function $F_k(.)$ whose i-th finite moment is $\beta_k^{(i)}$, i=1,2. The system is turned on when the number of priority-1 units waiting at the service facility exceeds θ, where θ is a nonnegative constant and the server is turned off when there are no units waiting for service. The value of θ that optimizes the total cost involving the start-up, shut-down and dormant costs as well as running and holding costs for undiscounted infinite horizon model is determined by using the longrun expected queue lengths, expected waiting times and other related characteristics. In the presence of start-up and shut-down costs, the policy of keeping the server in dormant state until the priority queue is built up to a certain level may considerably reduce the operating cost.

For any $t \geq 0$, let $U_k(t)$ (k=1,2) be the number of priority-k units in the system at time t. Any instant at which the server is turned off is a point of regeneration for the process $\{U_k(t), t \geq 0\}$, k=1,2. It is assumed that the server is turned off at the instant 0. The instant X is the next instant at which the server is turned off. Let a busy cycle be the time interval between two successive instants at which the server is turned off. The long run expected number of priority-k units in the system U_k is given as

$$U_k = E\left[\int_0^X U_k(s)\, ds\right]/E(X) \qquad (5.3.1)$$

where the integral in the numerator represents the total time spent by priority-k units in the system during one busy cycle and X represents the length of one busy cycle.

Let the probability distribution function G(x) be defined such that 1 and 0 are points of increase of G(x) with weights p_1 and p_2, where $p_i = \alpha_i / \Sigma \alpha_i$. Let $G^n(x)$ be the n-fold convolution of the probability distribution function G(x) with itself. Let us cosider the renewal function $N(x) = \sum_{n=1}^{\infty} G^n(x)$, $x \geq 0$, which is bounded to the function

$$N(x) = G(x) + \int_0^x G(x-y)\, dN(y), \quad x \geq 0 \qquad (5.3.2)$$

For any $s \geq 0$, let $M_k(s)$ be the number of priority-k units arriving in (0,s]. For any $x \geq 0$, let

$$S(x) = \inf[s/M_1(s) > x]$$

$$g_k(x) = M_k[S(x)]$$

$$W_k(x) = \int_0^{S(x)} U_k(s)ds, \quad k=1,2 \qquad (5.3.3)$$

The number of priority-k units in the system at the first instant at which the server is turned on is represented by $g_k(x)$, where $g_1(x) \geq 0$ and $g_2(x) \geq 0$ and the total time spent by priority-k units in the system upto that instant is represented by $W_k(x)$. For any $x \geq 0$, let

$$g(x) = g_1(x) + g_2(x)$$

$$c_k(x) = E[g_k(x)], \quad d_k(x) = E[g_k(x)\,(g_k(x)-1)]$$

$$w_k(x) = E[W_k(x)]$$

Further, let

$c_k(v) = 0$ for $v<0$ and

$a_1(x) = p_1+2p_1c_1(x-1)$ and

$a_2(x) = p_2+2p_2c_2x$ for $x \geq 0$

Then, for any $x \geq 0$,

$$c_k(x) = p_k((x+1)/p_1), \quad k=1,2$$

$$d_k(x) = a_k(x) + \int_0^x a_k(x-y)\, dN(y) - c_k(x), \quad k=1,2$$

Also, for any $x \geq 0$, let

$$c(x) = E[g(x)]$$

$$d(x) = E[g(x)\ (g(x)-1)]$$

$$e(x) = E[g_1(x)\ g_2(x)]$$

Then,

$$c(x) = (1+x)/p_1$$

$$d(x) = 2[(x+1)/p_1 - 1] + 2\int_0^x N(x-y)dN(y)$$

$$e(x) = (1/2)[d(x)-d_1(x)-d_2(x)] \qquad (5.3.4)$$

Let $c(v) = 0$ for $v<0$ and

$b_1(x) = (1/\alpha)c(x-1)$ and $b_2(x) = (1/\alpha)cx$ for $x \geq 0$

Then, for $x \geq 0$,

$$w_k(x) = b_k(x) + \int_0^x b_k(x-y)\ dN(y),\ k=1,2$$

Also,

$$b_1(\theta) = \theta(\theta+1)$$

$$b_2(\theta) = (\alpha_2^2/\alpha_1^2)(\theta+1)(\theta+2)$$

$$w_1(\theta) = (1/2\alpha_1)\theta(\theta+1)$$

$$w_2(\theta) = (\alpha_2/2\alpha_1^2)(\theta+1)(\theta+2)$$

$$c(\theta) = (\alpha_2/\alpha_1)\ (\theta+1)^2 \qquad (5.3.5)$$

Let $T(m,n)$ be the time elapsed from the start of a service when m priority-1 units and n priority-2 units are present in the system until the next instant at which the

system is empty. Let $t(m,n)=E[T(m,n)]$. Also, let $r_k(m,n)$ be the expected total time spent by priority-k units in the system during the time $T(m,n)$, k=1,2. Then, for m,n=0,1,2... and m+n>0,

$$t(m,n) = t_{b1}m+[n+\alpha_2 t_{b1}m]t_{b2}$$

$$r_1(m,n) = w_1 m+(1/2)t_{b1} m(m-1) + (\alpha_2 t_{b1}m+n) r_1(0,1)$$

$$r_2(m,n) = t_{b1}(1+\alpha_2 t_{b2})mn + (\alpha_2 t_{b1}m+n)r_2(0,1)$$
$$+(1/2)(\alpha_2+\alpha_2^2 t_{b2})[t_{b1}^{(2)}m+t_{b1}^{(2)}m(m-1)]$$
$$+(1/2)t_{b2} n(n-1)$$

where

$$t_{b1} = \beta_1/(1-\rho_1),\ t_{b1}^{(2)} = \beta_1^{(2)}/(1-\rho_1)^3,\ t_{b2} = \beta_2/(1-\rho)$$

$$w_1 = \beta_1/(1-\rho_1)+\alpha_1\beta_1^{(2)}/2(1-\rho_1)^2$$

$$r_1(0,1) = (1-\rho)^{-1}[\beta_1^{(2)}\alpha_1^{(2)}\beta_2/(2(1-\rho_1))+\beta_2^{(2)}\alpha_1/(2+\rho_1\beta_2)] \qquad (5.3.6)$$

$$r_2(0,1) = (1-\rho)^{-1}[\beta_1^{(2)}\alpha_1\rho_2/(2(1-\rho_1)(1-\rho))+\beta_2^{(2)}\alpha_2/(2(1-\rho))+$$
$$(1-\rho_1)\beta_2],$$

$$\rho_k = \alpha_k/\beta_k \text{ and } \rho = \sum_k \rho_k$$

The expected length of one busy cycle is given as

$$E(x) = (1/\alpha)(1+\theta)/p_1 + E[t(g_1(\theta),\ g_2(\theta)]$$
$$= ((1+\theta)p_1)/(\alpha(1-\rho))$$

The expected total time spent by the priority-k units in the system during one busy cycle is found to be equal to $w_k(\theta)+E[r_k(g_1(\theta),g_2(\theta))]$, k=1,2

From (5.3.1), the longrun expected number of priority-k

units U_k in the system shall be obtained as

$\alpha(1-\rho)[(1+\theta)/p_1]^{-1}[w_k(\theta+E\{r_k(g_1(\theta),g_2(\theta))\}]$, k=1,2

Using (5.3.5) and (5.3.6), U_1 and U_2 shall be expressed as

$$U_1 = H_1+((1-\rho)\theta)/(2(1-\rho_1)) \qquad (5.3.7)$$

$$U_2 = H_2+(\alpha_2\theta+2\alpha_2(1-\rho_1))/(2\alpha_1(1-\rho_1)) \qquad (5.3.8)$$

where

$H_k = \rho_k+\alpha_k[\alpha_1\beta_1^{(2)}+\alpha_2\beta_2^{(2)}]/[2(1-\rho_1)(1-\eta_k)]$,

k=1,2, $\eta_1=0$ and $\eta_2=\rho$

We proceed to consider a cost model involving the following costs. The dormant cost rate r_1 which is incurred when the server is dormant; the running cost rate r_2 which is incurred when the server is running; the start-up and shut-down cost rates R_1 and R_2 which are incurred whenever the server is activated and deactivated and the holding cost rates h_1 and h_2 which are paid as penalties for delaying priority-1 and priority-2 units in the system. While the increment (r_2-r_1) is incurred the fraction ρ of time when the server is busy, the cost (M_1+M_2) is incurred during each busy cycle. The holding costs h_1 and h_2 are proportional to the long run expected number of priority-1 and priority-2 units in the system. Thus the cost rate $C(\theta)$ is expressed as

$$\begin{aligned} C(\theta) = {} & r_1+(r_2-r_1)\rho + h_1[H_1+((1-\rho)\theta)/(2(1-\rho_1))] \\ & +h_2[H_2+(\alpha_2\theta+2\alpha_2(1-\rho_1))/(2\alpha_1(1-\rho_1))] \\ & +(M_1+M_2)(\alpha(1-\rho))/((1+\theta)p_1) \end{aligned} \qquad (5.3.9)$$

By setting $dC(\theta)/d\theta$ to zero and solving the resultant equation, the optimal value of θ is obtained as

$\theta^* = \psi^{\frac{1}{2}}-1$ or $-(1+\psi^{\frac{1}{2}})$, where

$$\psi = [2\alpha_1(1-\rho_1)(M_1+M_2)(\alpha(1-\rho))]/[p_1(h_1(1-\rho)\alpha_1+h_2\alpha_2)] \qquad (5.3.10)$$

The value of θ has to be an integer for specifying the optimal policy. If θ^* is not an integer, one of the integers surrounding θ^* is the best positive integer value of θ due to convexity of $C(\theta)$. The decision is taken by evaluating and comparing the cost rates for these surrounding integers.

Let us consider the case of changing β to β', where the random variable $\beta' = \lambda\beta$. The corresponding change in the running cost r_2 is $r_2' = r_2/\lambda$, while other costs remain the same. Also, $\rho^* = \alpha E(\lambda\mu) = \lambda\rho$. When $\lambda > 1$, β' is larger than β and this indicates a decrease in the capability of the server. When $\lambda < 1$, it represents an increase in the capability of the server. The optimal value of θ in this case is given by

$\theta^*(\lambda) = \delta^{\frac{1}{2}}-1$ or $-(1+\delta^{\frac{1}{2}})$, where

$$\delta = [2\alpha_1(1-\lambda\rho_1)(M_1+M_2)(\alpha(1-\lambda\rho))]/[p_1(h_1\alpha_1(1-\lambda\rho)+h_2\alpha_2)]$$

Thus, the cost rate for $\theta^*(\lambda)$ is obtained as

$$C(\lambda) = r_1+r_2\rho-r_1\lambda\rho+h_1[J_1+(1-\lambda\rho)\Theta^*(\lambda))/(2(1-\lambda\rho_1))]$$
$$+h_2[J_2+(\alpha_2\Theta^*(\lambda)+2\alpha_2(1-\lambda\rho_1))/(2\alpha_1(1-\lambda\rho_1))]$$
$$+(M_1+M_2)(\alpha(1-\lambda\rho))/[(1+\Theta^*(\lambda))p_1]$$

where

$$J_k=\lambda\rho_k+\alpha_k[\alpha_1\beta_1^{(2)}+\alpha_2\beta_2^{(2)}]/[2(1-\lambda\rho_1)(1-\pi_k)],$$

$$k=1,2; \quad \pi_1=0 \text{ and } \pi_2=\lambda \qquad (5.3.11)$$

Since it is difficult to get the solution of the equation $\partial C(\lambda)/\partial\lambda=0$ in closed form, the condition for which it is desirable to slow down the server shall be obtained when the initial condition is $\lambda=1$ as $dC(\lambda)/d\lambda|_{\lambda=1} < 0$

After simplification, the condition for slowing down the server is obtained as

$$h_1[4\rho_1(1-\rho_1)^2+\alpha_1\rho_1(\alpha_1\beta_1^{(2)}+\alpha_2\beta_2^{(2)})$$
$$+2(1-\rho_1)((1-\rho)\Theta^{*\prime}(1)-\Theta^*(1))+2(1-\rho)\rho_1\Theta^*(1)]/(4(1-\rho_1)^2)$$
$$+h_2[2(1-\rho)^2(1-\rho_1)^2\rho_2+\alpha(\alpha_1\beta_1^{(2)}+\alpha_2\beta_2^{(2)})(\rho(1-\rho_1)+\rho_1(1-\rho))]/$$
$$[2(1-\rho_1)^2(1-\rho)^2]-r_1\rho<(M_1+M_2)[p_1\alpha(1+\Theta^*(1))$$
$$+\alpha p_1(1-\rho)\Theta^{*\prime}(1)]/[p_1^2(1+\Theta^*(1))^2]$$

The optimal value of σ under the policy in which the server is turned on if the number of priority-2 units waiting for service exceeds a non-negative integer σ shall be determined by following the earlier approach. The long run expected queue lengths in this case shall be obtained as

$$U_1 = H_1+[\alpha_1(1-\rho)(\sigma+2)]/[2\alpha_2(1-\rho_1)]$$

$$U_2 = H_2+(2\rho_1+\sigma)/(2(1-\rho_1))$$

The cost rate $C(\sigma)$ is obtained as

$$C(\sigma) = r_1+(r_2-r_1)\rho + h_1[H_1+\{\alpha_1(\sigma+2)(1-\rho)\}/(2\alpha_2(1-\rho_1))]$$
$$+h_2(H_2+(2\rho_1+\sigma)/(2(1-\rho_1))+(M_1+M_2)[\alpha(1-\rho)]/(\sigma+1)$$

The optimal value of σ shall be obtained by setting $dC(\sigma)/d\sigma$ to zero as

$\sigma^* = \mu^{\frac{1}{2}}-1$ or $-(1+\mu^{\frac{1}{2}})$, where

$$\mu = [2\alpha_2(1-\rho_1)(M_1+M_2)(\alpha(1-\rho))]/[h_1\alpha_1(1-\rho)+h_2\alpha_2]$$

If σ^* is not an integer, one of the integers surrounding σ^* is the best positive integer value of n, which shall be found by evaluating and comparing the cost rates for these surrounding integers.