Pricing surplus server capacity for mean waiting time sensitive customers

Sudhir K. Sinha, N. Rangaraj and N. Hemachandra

Industrial Engineering and Operations Research,

Indian Institute of Technology Bombay, Mumbai 400076, India

September 6, 2008

Abstract

Resources including various assets of supply chains, face random demand over time and can be shared by others. We consider an operational setting where a resource is shared by two different classes of customers, the primary class (existing customers) and the secondary class (new firms) customers, under a service level based pricing contract with the secondary class customers which specifies the unit price of service as well as the quality of service offered. We assume that the Poisson arrival rate of secondary class of customers depends linearly on the unit price of service as well as on the service level offered. In our model, we use delay dependent priority scheme for queue management and stationary mean expected time as a quality of service measure. Given an existing service level based contract between the service provider and the primary class of customers, we analyze the impact of inclusion of secondary class of customers has on the system utilization and service level of the existing customers. We study the joint problem of optimally pricing and operation of the resource with the inclusion of the secondary class of customers, while continuing to offer a pre-specified quality of service to primary class of customers. This non-convex constrained optimization problem has two pricing (contract) decision variables and two operational decision variables. While the two decision variables, the unit price of service and quality of service level offered constitute the pricing parameters, the allowable rate of secondary class customers and a parameter capturing the relative delay dependent queue priority form the two operational decision variables. We observe that in our model, we can first find the optimal operating parameters and then use them to find the optimal contract (pricing) parameters. This follows from separability property of linear demand function. We search exhaustively for Karush-Kuhn-Tucker points of the optimization problem to obtain the global optimum point. We propose an algorithm that finds these optimal parameters in closed form expressions for various possible values of input parameters. We also study in detail the structure and the non-linear nature of these optimal decisions along with their sensitivity to various input parameters.

The study has implications in settings where a new firm enters into business requiring high infrastructural set up cost and is willing to use the infrastructure of an existing firm after entering into a pricing contract. An example is the entry of private firms into the inland rail container movement business in India.

Key words: Queueing, quality of service, dynamic priority schemes, linear demand function, non-convex optimization

1 Introduction

In addition to cost minimization, guaranteeing assured levels of service has been a dominant concern while operating resources. In practice, it is possible that resources remain under-utilized because of the random nature of demand and usage. Owners of such resources may want to share the existing resources with others, including new firms, requiring such resources. In such a scenario, the resource will be used by two different classes or types of customers; the primary class customers (existing customers) and secondary class customers (customers of other or new firms). Queueing systems are natural models for resource allocation to customers who arrive over time. In this paper, we propose a priority queue based model for the optimal use of excess capacity of a resource when customers of both classes arrive over time and these customers will be offered pre-specified Quality of Service (QoS) level guarantees. In particular, we consider the issue of optimal pricing of excess capacity of a server for an independent Poisson stream of secondary class customers whose arrival rate is sensitive to both the offered mean waiting time (QoS level) as well as to the price charged per customer while simultaneously ensuring that the mean waiting time (QoS level) of the primary class customers is less than a pre-specified level.

We assume that the resource owner has a long term agreement with the primary class customers which specifies a QoS level to the primary class customers. The agreement assumes a Poisson demand for the primary class customers. The inclusion of secondary class customers into the system increases the traffic intensity at the resource, which in turn, affects not only the system utilization but also the effective service level offered to the primary class customers. Therefore, one needs to control the arrival rate of the secondary class customers into the system. We assume that the demand of the secondary class customers depends not only on the price charged but also on the assured service level. Such a pricing scheme will influence the secondary class customers' demand and hence it can be viewed as a mechanism to control the traffic intensity at the resource.

The customer's service level depends on the queueing discipline at the resource. A queue discipline used will therefore affect the secondary class customers' arrival rate whose demand is sensitive to unit admission price as well as to the assured service level. That, in turn, also affects the effective service level offered to the primary class customers.

Thus, the queue discipline employed can be viewed as another mechanism to control the traffic intensity at the resource. The simplest queueing discipline is first come first serve (FCFS). But, FCFS queueing discipline does not provide differentiated service level. A differentiated service level between customers of different classes can be achieved using priority queue management discipline. Such a priority scheme can be either static or dynamic. The static priority scheme may cause long delays to the jobs of low priority customers. A dynamic priority scheme eliminates the disadvantages of FCFS and static priority disciplines as jobs for service are now selected based on their actual waiting times as well as their priority class. In our model for resource sharing we use a delay dependent priority queue management scheme originally proposed by Kleinrock (1964).

In this paper, we restrict ourselves to the case when QoS levels of a class are measured in terms of the stationary expected waiting time in queue of customers of that class. An assured service level for a class implies that the stationary expected waiting time of customers in queue of that class will be less than or equal to the assured mean waiting time. We also assume that the potential mean arrival rate of the secondary class customers is a linear function of unit admission price and assured service level. The resource owner aims to select a pair of operating parameters, a queue discipline management parameter characterizing the dynamic priority policy and an appropriate arrival rate of the secondary class customers along with a suitable pair of pricing parameters, i.e., unit admission price and assured mean waiting time for the secondary class customers, that will maximize its expected revenue from the inclusion of secondary class customers while ensuring the prevailing mean waiting time level to the primary class customers. Such a constrained resource sharing problem can be viewed as a design of a QoS level based contract that the resource owner wants to enter with the secondary customers. Given that secondary customers' Poisson arrival rate is linear in unit admission price and assured service (mean waiting time in queue) level, the resource owner would like to quote optimal values for these two quantities that also ensures pre-assigned QoS (mean waiting time in queue) level to the primary customers. The secondary customers' market will offer an additional steady Poisson demand for the resource owner while availing a certain QoS (mean waiting time) that is specified in the contract by the resource owner. The resource owner will employ a dynamic priority management scheme to meet these QoS levels of both the classes of customers. The stationary waiting time of a class in the queueing model can be interpreted as the sample path based customer average of waiting times of members of that class in a regenerative system like ours (Wolff 1989). This suggests a practical way to implement such a long term contract that the resource owner may want to enter into with the secondary class customers.

The assumption on the nature of demand function and the definition of customers' service level help us to reformulate the resource owner's constrained problem. The decision variables of the reformulated problem are the operating parameters, the arrival rate of the secondary class customers and the queue discipline management parameter, while the suitable pair of pricing (contract) parameters, i.e., the unit admission price and the assured mean waiting time for the secondary class customers, are derived using the values of the decision variables at optimality of the reformulated problem. By choosing various possible values of the Lagrange variables in the Karush-Kuhn-Tucker first order necessary conditions of this reformulated optimization problem, we exhaustively search for its Karush-Kuhn-Tucker points. One of the decision variables, the queue discipline management parameter, can take the value of infinity (corresponding to head-of-the-line static high priority to the secondary class customers) and hence constitutes a valid decision in our optimization problem. Therefore, we also separately analyze the constrained optimization problem with arrival rate of secondary class customers as a single decision variable, by setting the queue discipline management parameter as infinity. We next compare the optimal solutions of these two optimization problems to obtain the global optimal solution of the original constrained resource sharing problem. We identify its global optimal solution for all possible values of input parameters, except for one finite interval of S_p, the prevailing QoS level of primary class customers. Based on our analysis and numerical experimentation, we conjecture an optimal solution for this finite interval of S_p also. This leads to an algorithm that terminates in finite steps with closed form expressions for optimal values of both the pairs of operating parameters and pricing parameters.

A consequence of the use of the linear demand function in the context of sharing a resource over time is that the joint problem of optimal pricing and operation of excess capacity can be separated; one can find optimal operating variables first and then use them to find the optimal pricing parameters. Also, it turns out that the optimal decision variables remain insensitive to the price sensitivity coefficient of the demand function. One of our findings is that there exits an interval for the ratio of co-efficients of linear demand function such that it is beneficial for the resource owner to offer static high priority to the secondary class customers. For values of the above ratio to the right of this interval, it may be optimal to use dynamic priority scheme. In such a case, a part of the feasible values of S_p will have three intervals and in these intervals exactly one of the two operating parameters, either the optimal arrival rate of the secondary class of customers, λ_s^*, or the parameter corresponding to optimal dynamic priority queue management, β^*, remains constant. But, the optimum contract (pricing) parameters θ^* and S_s^* are different non-linear functions in the different intervals of S_p.

The sensitivity analysis of optimal parameters with respect to demand coefficients shows that the increase in the maximum attainable demand rate of the secondary class customers means that a delay dependent priority queue discipline needs to be used as part of optimal policy over a wider range of S_p values. The increase in service level sensitivity coefficient of secondary class customer leads to the use of the delay dependent priority queue discipline for a smaller range of S_p values.

This work is motivated by the recent opening up of the inland rail container movement in India to both the private and public sector players, which till recently was solely managed by Container Corporation of India Ltd (Concor), a public firm. The interested companies have to arrange for a rail-linked inland container depot (ICD). Due to the high infrastructural set up cost involved, new firms may be seeking to lease some resources like ICDs from Concor (presently the only one to possess a rail-linked ICD) in the initial years of operations. In a shared ICD, the existing customers of Concor are the primary class customers whereas customers of the new firms constitute the secondary class customers (Sinha *et al.* 2008). The framework we consider is optimal pricing of surplus capacity in a general commitment based resource sharing model and can be potentially relevant in many settings, e.g., a in-house manufacturing unit of a firm utilizing its excess capacity to cater an outside firm's demands, a third party logistics service provider serving multiple customers. Other type of situations can be communication networks providing service to different classes of customers. A contemporary example could be the determination of charges a mobile telephony service provider can use while providing roaming services to customers of another service provider.

Similar studies of Palaka *et al.* (1998), Pekgun *et al.* (2008), Ray and Jewkes (2004) and So and Song (1998) determine an optimum pair of price and quoted lead-time for customers sensitive to price as well as quoted lead time. The quoted lead time is identical to assured service level. These studies assume a single class of customers and employ FCFS queueing discipline. Palaka *et al.* (1998) and Pekgun *et al.* (2008) model customer demand as a linear function of price and quoted lead time. The linear demand nature will mean that secondary customers view price and service level as substitutes (Palaka *et al.* 1998). Such demand functions also exhibit nice elasticity properties that we summarize later. Palaka *et al.* (1998) model the system as $M/M/1$ queue and consider that the firm incurs congestion cost as well as pays lateness penalties. They find the optimal decisions of the resulting profit maximization problem and study the impact of varying parameters values on those optimal decisions. They also examine a situation in which it is possible for the firm to expand capacity marginally. Pekgun *et al.* (2008) consider a firm where pricing and lead time decisions are made by two independent functions, marketing and production, respectively. They model the firm's operations as a $M/M/1$ queue and the sequence of decisions as a Stackelberg game. They show the inefficiencies resulting from decentralized decision making and present a coordination scheme to overcome those inefficiencies. Their focus is on the desirability of coordination issues in this setting.

Ray and Jewkes (2004) extend the linear customer demand model by assuming that price itself is function of lead time. They assume that the firm can reduce the lead time by investment in capacity. They first determine the profit maximizing optimal policy and thereafter investigate the behaviour of the optimal policy under various changes of the system parameters. They specifically present the conditions under which overlooking price and lead time dependence will lead to a sub-optimal decision. They also extend the model by incorporating economies of scale to unit operating cost. In contrast, So and Song (1998) consider log-linear Cobb-Douglas demand function to reflect customer's sensitivity to price and lead time and model a service facility as a $G/G/1$ queue. They determine the optimal price, lead (delivery) time quote and short-term capacity expansion level which maximizes the average net profit while maintaining a predetermined level of delivery reliability. We restrict ourselves to a linear demand model. We advocate the use of dynamic queue management schemes as part of service level based pricing in multi-class queueing models.

Hall *et al.* (2002) study a similar setting where a resource is shared by two different classes of customers. They assume FCFS queue discipline at resource and also assume that the demand is sensitive to just unit price. They focus on dynamic pricing policies which depend on the production system (queue) status. They demonstrate the properties of the optimal policies and show that a policy of uniform pricing up to cutoff state is superior according to a certain performance/complexity ratio measure. We, in contrast, focus on static pricing scheme with dynamic queue discipline management.

We present the details of the operational setting and the optimal constrained resource sharing problem in Section 2. Analysis of this optimization problem are given in Section 3. Based on it, we present in Section 4 the algorithm to select the optimal contract parameters for secondary class customers and optimal parameters to operate the resource and also illustrate the algorithm by several numerical examples that correspond to various possible cases of the input data parameters of the model. Further, we present details of our understanding of the model and sensitivity analysis of its optimal solutions in Section 5. A preliminary version of the algorithm along with a numerical example are presented in Sinha *et al.* (2008). In the present paper, we present detailed arguments that lead to the algorithm along with an extensive sensitivity analysis.

2 A queueing model for a shared resource

Let λ_p and λ_s be independent Poisson arrival rates of the customers of the primary and secondary classes respectively. As the service requirements of the primary and secondary class customers are identical in nature, we assume that the service times, i.e., time taken by the resource to complete a job irrespective of customers' class, are independent and identically distributed random variables with mean $1/\mu$ and variance σ^2. Further, the queue discipline employed at the resource is head-on-line (non-preemptive) delay dependent priority scheme. A schematic view of the system is shown in Figure 1.

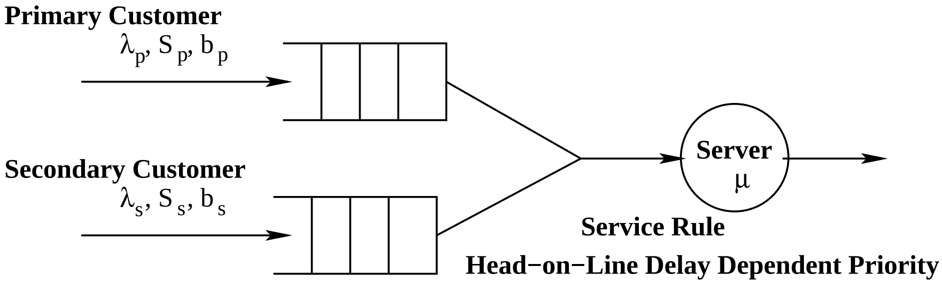


Figure 1: Schematic view of the shared resource

The delay dependent priority scheme is an example of dynamic priority scheme. A queueing system with delay dependent priority scheme was first studied by Kleinrock (1964). It consists of P priority classes associated with a set of variable parameters $\{b_p\}_1^P$, where $0 \leq b_1 \leq b_2 \leq \ldots \leq b_P$. The instantaneous priority at time t of a class p job that arrived at time T_p is given by $q_p(t) = (t - T_p)b_p$. After a service completion, the server chooses the next job with highest instantaneous priority $q_p(\cdot)$ from all available jobs. If there is a tie for the highest instantaneous priority, then it is broken by using FCFS rule. Here, a higher priority job gains priority at faster rate than lower priority jobs. The steady state expected waiting time in queue for a class p job in M/G/1 head-on-line delay dependent queue is given by the following recursion (Kleinrock 1964, Kanet 1982):

$$W_p = \frac{\dfrac{W_0}{1-\rho} - \displaystyle\sum_{i=1}^{p-1} \rho_i W_i \left(1 - \frac{b_i}{b_p}\right)}{1 - \displaystyle\sum_{i=p+1}^{P} \rho_i \left(1 - \frac{b_p}{b_i}\right)}, \qquad p \in \{1, 2, \ldots, P\} \tag{1}$$

where $\rho_i = \dfrac{\lambda_i}{\mu_i}$, $\rho = \displaystyle\sum_{p=1}^{P} \rho_i$, $W_0 = \displaystyle\sum_{p=1}^{P} \frac{\lambda_p}{2}\left(\sigma_p^2 + \frac{1}{\mu_p^2}\right)$ and $0 \leq \rho < 1$. We note that the queue parameters $\{b_p\}_1^P$ only appear as ratios b_p/b_{p+1} in the expression for W_p. Also, the conservation law for M/G/1 system with non-preemptive work-conserving queueing discipline (a system in which work is neither created nor destroyed within the system) (Kleinrock 1976) states that

$$\sum_{p=1}^{P} \rho_p W_p = \begin{cases} \dfrac{\rho W_0}{1-\rho} & \rho < 1 \\ \infty & \rho \geq 1 \end{cases}$$

Let b_p and b_s be the associated parameters of the primary and secondary class customers respectively in our system. Alternatively, we define relative priority queue discipline management parameter β as a ratio of the associated queue parameters to the primary and the secondary class customers, i.e., $\beta := b_s/b_p$. The selection of the relative priority control parameter β defines different regimes of the delay dependent priority queue.

- $\beta < 1$ ($b_s < b_p$). This implies that the primary class customers get higher priority than the secondary class customers. When β approaches zero, the queuing system becomes equivalent to a static priority queue with high priority to primary class customers.
- $\beta = 1$ ($b_s = b_p$). This implies that both classes of customers get equal priorities. Thus, the queueing discipline is FCFS.
- $\beta > 1$ ($b_s > b_p$). This implies that the secondary class customers get higher priority than the primary class customers. When β approaches infinity the queuing system becomes equivalent to a static priority queue with high priority to secondary class customers.

Figure 2 summarizes the relationship between the control parameter β and priorities of the primary and secondary class customers.

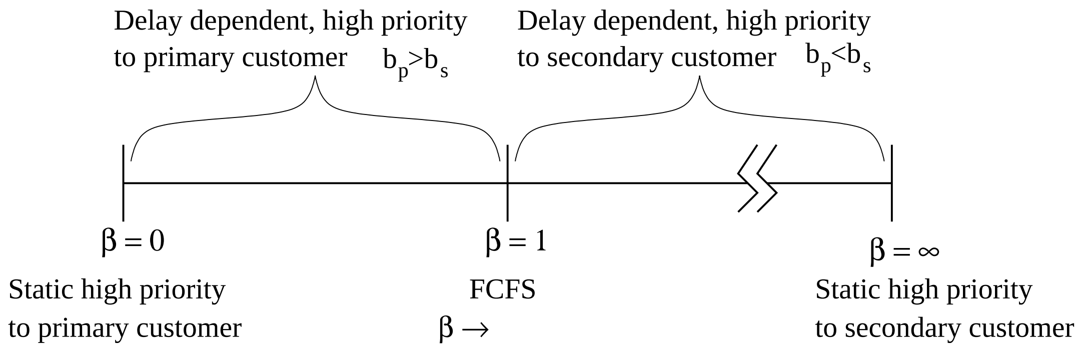


Figure 2: Effect of queue discipline management parameter on queue discipline

Recall that we assume that the prevailing agreement between the resource owner and the primary class customers results in a Poisson arrival rate λ_p of the primary class customers. In the absence of the secondary class customers, it is assumed that the resource owner is able to fulfill the assured mean waiting time requirement of the primary class customers, given by S_p, and still the facility is under-utilized. This assumption holds if and only if $S_p \geq \hat{S}_p$, where $\hat{S}_p$ is the expected waiting time in queue for the primary class customers when the resource is *dedicated* to the primary class customers. In the M/G/1 setting, $\hat{S}_p = \frac{\lambda_p \psi}{\mu(\mu - \lambda_p)}$ where $\psi = \left[1 + \sigma^2 \mu^2\right]/2$.

Let $\Lambda_s(\theta, S_s)$ be the corresponding potential Poisson arrival rate of the secondary class customers with a charged unit admission price of θ per completed job and assured service level of S_s. We assume that this rate is a linear function of the unit admission price and assured service level, i.e.,

$$\Lambda_s(\theta, S_s) = a - b\theta - cS_s \tag{2}$$

for some given constants $a, b, c > 0$. The coefficients a, b and c represent the maximum attainable mean arrival rate (market potential), price sensitivity and service level sensitivity respectively. We have assumed linear demand function because it is convenient and it exhibits the following desirable properties (Palaka *et al.* 1998).

1. The demand function is separable in price and assured service level and thus makes price and service level as substitutes.
2. The price elasticity of demand, $-b\theta/(a - b\theta - cS_s)$, increases with increase in θ and S_s. This results in higher price elasticity of demand at higher unit price θ and higher service level S_s. Thus, customers are more sensitive to high prices when they wait for longer periods of time in the queue. A similar property is valid regarding service level elasticity.

As the arrival rate of the primary class customers remains fixed at λ_p, the expected waiting times of the customers depend only on the mean arrival rate of the secondary class customers λ_s and relative queue discipline management parameter β. Let $W_p(\lambda_s, \beta)$ and $W_s(\lambda_s, \beta)$ be the expected waiting times in the queue for the primary and the secondary class customers respectively.

The resource owner aims to select an appropriate arrival rate of the secondary class customers λ_s, a suitable pair of pricing parameters θ and S_s for the secondary class customers and a queue discipline management parameter β that will maximize its expected revenue from the inclusion of secondary class customers while ensuring the prevailing mean waiting time to the primary class customers. The resulting constrained resource sharing problem of the resource owner is as follows

$$\textbf{P0}: \quad \max_{\lambda_s, \theta, S_s, \beta} \theta\lambda_s \tag{3}$$

$$\text{Subject to:} \quad W_p(\lambda_s, \beta) \leq S_p \tag{4}$$

$$S_s \geq W_s(\lambda_s, \beta) \tag{5}$$

$$\lambda_s \leq \mu - \lambda_p \tag{6}$$

$$\lambda_s \leq a - b\theta - cS_s \tag{7}$$

$$\lambda_s, \theta, S_s, \beta \geq 0. \tag{8}$$

Here, constraint (4) ensures that the resource owner does not violate the prevailing service level commitment to the primary class customers while sharing the resource. Constraint (5) restricts the resource owner to offer a service level commitment to secondary class customers within system capability. Constraint (6) sets a restriction on maximum permissible mean arrival rate of secondary class customers based on processing capability of the system, i.e., it avoids instability of the multi-class queue. Later we show that this constraint indeed remains non-binding at the optimum and ensures stability of the multi-class queue. Constraint (7) ensures that the mean arrival rate of secondary class customers should not exceed the demand generated by charged price θ and offered service level S_s. The last constraint captures the non-negativity of the mean arrival rate of secondary class customers λ_s, price θ, assured service level S_s and queue discipline management parameter β.

Constraint (7) will hold as an equality in an optimal solution given that the demand is a separable function of both price and assured service level (Palaka *et al.* 1998). Next, we claim that constraint (5) will also hold as an equality in an optimal solution. For, suppose that the optimal solution of the optimization problem is given by mean arrival rate of secondary class customers λ_s^*, price θ^*, assured service level S_s^* and queue discipline management parameter β^* that satisfy $S_s^* > W_s(\lambda_s^*, \beta^*)$. We already know that constraint (7) will hold as an equality in an optimal solution. Therefore, the objective function can be rewritten as $\frac{1}{b}\left[a\lambda_s - \lambda_s^2 - c\lambda_s S_s\right]$. Since the objective function is a decreasing function of S_s, an assured service level S_s' such that $S_s^* > S_s' \geq W_s(\lambda_s^*, \beta^*)$ will increase the earned revenue of the resource owner. Therefore, the constraint (5) must hold as an equality in the optimal solution.

In view of the fact that the above constraints are binding at optimality, the resource sharing problem of the facility owner, (P0) can be rewritten as

$$\textbf{P1}: \qquad \max_{\lambda_s, \beta} \ \frac{1}{b}\left[a\lambda_s - \lambda_s^2 - c\lambda_s W_s(\lambda_s, \beta)\right] \tag{9}$$

$$\text{Subject to:} \qquad W_p(\lambda_s, \beta) \leq S_p \tag{10}$$

$$\lambda_s \leq \mu - \lambda_p \tag{11}$$

$$\lambda_s, \beta \geq 0. \tag{12}$$

Once the optimal mean arrival rate of secondary class customers λ_s^* and queue discipline management parameter β^* is known, the optimal price θ^* and assured service level S_s^* is obtained using equalities $\lambda_s^* = a - b\theta^* - cS_s^*$ and $S_s^* = W_s(\lambda_s^*, \beta^*)$. The objective function (9) indicates that the optimal choices of λ_s and β remain *insensitive* to price sensitivity co-efficient b of the secondary class customers.

Further, we claim that the constraint (11) should remain *non-binding* at the optimal point, i.e., the constraint (11) will hold with strict inequality at optimality, $\lambda_s^* < \mu - \lambda_p$. To arrive at a contradiction, let us assume that the constraint (11) is binding at the optimal point, i.e., $\lambda_s^* = \mu - \lambda_p$. We note that as S_p is finite, $W_p(\lambda_s^*, \beta^*)$ will be finite at optimality. Suppose $S_s > \frac{a - b\theta}{c}$; then, the demand function (2) results in $\lambda_s < 0$. Thus, $\lambda_s \geq 0$ if and only if $S_s \leq \hat{S}_s(\theta)$ where $\hat{S}_s(\theta) = \frac{a - b\theta}{c}$. As $\hat{S}_s(\theta)$ is a finite number, the optimum S_s^* will remain finite. Also, at optimality $S_s^* = W_s(\lambda_s^*, \beta^*)$. Therefore, $W_s(\lambda_s^*, \beta^*)$ should take a finite value at optimality. When $\lambda_p + \lambda_s^* = \mu$, it is not possible to have finite $W_p(\lambda_s^*, \beta)$ and $W_s(\lambda_s^*, \beta)$ simultaneously for any feasible β. This contradicts the initial assumption that the constraint (11) is binding at optimality, i.e., $\lambda_s^* = \mu - \lambda_p$. Hence, the constraint (11) will remain *non-binding* at the optimal solution.

3 Optimal pricing and operation of the resource sharing model

Using recursion (1), the expected waiting times in queue for the primary and the secondary class customers are given by

$$W_p(\lambda_s, \beta) = \frac{\lambda\psi\left[\mu - \lambda\left[1-\beta\right]\right]}{\mu\left[\mu-\lambda\right]\left[\mu - \lambda_p\left[1-\beta\right]\right]}\mathbb{1}_{\{\beta\le 1\}} + \frac{\lambda\psi}{\left[\mu-\lambda\right]\left[\mu-\lambda_s\left[1-\frac{1}{\beta}\right]\right]}\mathbb{1}_{\{\beta>1\}} \tag{13}$$

$$W_s(\lambda_s, \beta) = \frac{\lambda\psi}{\left[\mu-\lambda\right]\left[\mu - \lambda_p\left[1-\beta\right]\right]}\mathbb{1}_{\{\beta\le 1\}} + \frac{\lambda\psi\left[\mu-\lambda\left[1-\frac{1}{\beta}\right]\right]}{\mu\left[\mu-\lambda\right]\left[\mu-\lambda_s\left[1-\frac{1}{\beta}\right]\right]}\mathbb{1}_{\{\beta>1\}} \tag{14}$$

where $\lambda = \lambda_p + \lambda_s$, $\psi = \left[1+\sigma^2\mu^2\right]/2$ and $\mathbb{1}_{\{.\}}$ denotes the indicator function which is equal to 1 if the statement between braces is true and 0 otherwise. We note that the objective function and the constraint (10) of the resource sharing problem P1 are defined differently in the regions corresponding to $\beta \le 1$ and $\beta > 1$. This aspect distinguishes the optimization problem P1 from a classical optimization problem.

First, we note some useful properties of $W_p(\lambda_s, \beta)$ and $W_s(\lambda_s, \beta)$ whose details are given in Appendix. Next, we show that the above optimization problem is a non-convex problem.

1. $W_p(\lambda_s, \beta)$ and $W_s(\lambda_s, \beta)$ are increasing convex function of λ_s in the interval $[0, \mu - \lambda_p)$.
2. $W_p(\lambda_s, \beta)$ is an increasing concave function of $\beta \ge 0$ whereas $W_s(\lambda_s, \beta)$ is a decreasing convex function of $\beta \ge 0$.
3. $W_p(\lambda_s, \beta)$ is neither a convex nor a concave function of (λ_s, β) where $\lambda_s \in [0, \mu - \lambda_p)$ and $\beta \ge 0$. Also, $W_p(\lambda_s, \beta)$ is not quasi-convex function of (λ_s, β); a numerical example is given below.
4. $\lambda_s W_s(\lambda_s, \beta)$ is neither a convex nor a concave function of (λ_s, β) where $\lambda_s \in [0, \mu - \lambda_p)$ and $\beta \ge 0$.

We demonstrate below by a numerical example that the $W_p(\lambda_s, \beta)$ is also not a quasi-convex function of (λ_s, β) where $\lambda_s \in [0, \mu - \lambda_p)$ and $\beta \ge 0$. Let us assume that $\lambda_p = 8$, $\mu = 10$ and $\sigma = 0.1$. We note that $W_p(1.5, 0) = 0.475$, $W_p(0.5, 1) = 0.567$, $W_p(1, 0.5) = 0.825$ and

$$W_p(0.5(1.5, 0) + 0.5(0.5, 1)) = W_p(1, 0.5) > \max\left\{W_p(1.5, 0), W_p(0.5, 1)\right\}.$$

The above inequality violates the necessary condition for a function to be a quasi-convex function (Bazaraa *et al.* 1993); hence, $W_p(\lambda_s, \beta)$ is not a quasi-convex function over $\lambda_s \in [0, \mu - \lambda_p)$ and $\beta \geq 0$. This means that, because constraint (10), the feasible region will be non-convex. Also, the Hessian matrix of $\lambda_s W_s(\lambda_s, \beta)$ at $(\lambda_s = 0.1, \beta = 0.5)$ is

$$\begin{pmatrix} 0.972 & -1.009 \\ -1.009 & 0.253 \end{pmatrix}.$$

The eigenvalues of this Hessian matrix are 1.684 and -0.459. This implies that $\lambda_s W_s(\lambda_s, \beta)$,i.e., third term of the objective function, is neither a convex nor a concave function of (λ_s, β). Hence, optimization problem P1 is a non-convex constrained optimization problem.

The Lagrangian function corresponding to the non-linear programming (NLP) problem P1 can be expressed as

$$L_1(\lambda_s, \beta, u_1, u_2, u_3) = \frac{1}{b}\left[a\lambda_s - \lambda_s^2 - c\lambda_s W_s(\lambda_s, \beta)\right] + u_1\left[W_p(\lambda_s, \beta) - S_p\right] + u_2\lambda_s + u_3\beta \quad (15)$$

where u_1, u_2 and u_3 are the Lagrangian multipliers. The optimum value of the vector $(\lambda_s, \beta, u_1, u_2, u_3)$ should satisfy the Karush-Kuhn-Tucker first order necessary conditions. These are given as follows (Bazaraa *et al.* 1993):

$$a - 2\lambda_s - c\left[W_s + \lambda_s \frac{\partial W_s}{\partial \lambda_s}\right] + bu_1 \frac{\partial W_p}{\partial \lambda_s} + bu_2 = 0 \quad (16)$$

$$-c\lambda_s \frac{\partial W_s}{\partial \beta} + bu_1 \frac{\partial W_p}{\partial \beta} + bu_3 = 0 \quad (17)$$

$$u_1\left[W_p - S_p\right] = 0 \quad (18)$$

$$u_2\lambda_s = 0 \quad (19)$$

$$u_3\beta = 0 \quad (20)$$

$$W_p \leq S_p \text{ and } \lambda_s < \mu - \lambda_p \quad (21)$$

$$u_1 \leq 0; \lambda_s, \beta, u_2, u_3 \geq 0. \quad (22)$$

A Karush-Kuhn-Tucker (KKT) point is defined by a specific vector $(\lambda_s, \beta, u_1, u_2, u_3)$ that satisfies the conditions (16)-(22). If the KKT point also satisfies the second order sufficient conditions then it can be either a local or a global optimum point of the NLP P1. We note that if the Lagrangian multiplier u_2 is such that $u_2 > 0$, then the KKT condition (19) is satisfied if and only if $\lambda_s = 0$, in which case objective function value is zero. As the objective of the resource owner is to earn a strict positive revenue, we ignore values of $u_2 > 0$ in further analysis and assume throughout that $u_2 = 0$. The analysis below exhaustively searches for all possible KKT points of the optimization problem P1 where $u_2 = 0$, i.e., assigns specific values to the remaining four unknown elements of a possible KKT point.

First, we investigate the KKT conditions (17), (20) and (22) in detail. Let us assume that $\beta > 0$ at optimum. If $\beta > 0$, then the KKT condition (20) is satisfied iff $u_3 = 0$. Given $u_3 = 0$, the KKT condition (17), using work conservation relation, results in

$$u_1 = \frac{c\lambda_s \frac{\partial W_s}{\partial \beta}}{b\frac{\partial W_p}{\partial \beta}} = -\frac{c\lambda_p}{b}. \tag{23}$$

Next, let us consider that $\beta = 0$ at optimum. The simplification of the KKT condition (17) at $\beta = 0$ results in

$$u_3 = -\frac{\lambda_s \lambda \psi [c\lambda_p + bu_1]}{b[\mu - \lambda][\mu - \lambda_p]^2}. \tag{24}$$

We note that $u_3 \geq 0$ iff $u_1 \leq -\frac{c\lambda_p}{b}$ given that $0 < \lambda_s < \mu - \lambda_p$ and $\lambda_p \geq 0$. In particular, $u_3 = 0$ at $u_1 = -\frac{c\lambda_p}{b}$. Thus, the KKT conditions (17), (20) and (22) are satisfied if and only if one of the following hold true:

C1: $u_1 = -\frac{c\lambda_p}{b}$, $u_3 = 0$ and $\beta \geq 0$.

C2: $u_1 < -\frac{c\lambda_p}{b}$, $u_3 = -\dfrac{\lambda_s \lambda \psi [c\lambda_p + bu_1]}{b[\mu - \lambda][\mu - \lambda_p]^2}$ and $\beta = 0$.

The above investigation explicitly assigns specific values to two unknown elements of a possible KKT point. The remaining two unknown elements of a possible KKT point is obtained by solving the Equations (16) and (18) (note that $u_2 = 0$ at optimum).

Next, we investigate the KKT condition (18) considering the fact that the constraint (10) can be either binding or non-binding at optimum. If the constraint (10) is binding at the optimum, then the KKT condition (18) gets automatically satisfied irrespective of the value of u_1. On the other hand, if the constraint (10) is non-binding at optimum, then the KKT condition (18) is satisfied if and only if $u_1 = 0$. But, we note from conditions C1-C2 that u_1 should be less than or equal to $-\frac{c\lambda_p}{b}$ to satisfy (17), (20) and (22). As $\frac{c\lambda_p}{b} \neq 0$, the KKT conditions (17), (18), (20) and (22) are not satisfied simultaneously at $u_1 = 0$. This implies that it is not possible to have a KKT point with the constraint (10) non-binding. Therefore, the constraint (10) will always be binding at optimum.

Further, we note that the classical optimization theory and thereby the KKT first order necessary conditions inherently assumes finite values of the decision variables. The above analysis which results in a particular value of Lagrangian multipliers also considers that both λ_s and β are finite. But, $\beta = \infty$ is a valid decision in our optimization problem. Therefore, we separately analyze the resulting one-dimensional optimization problem (by setting $\beta = \infty$ in the NLP P1) in Section 3.2. In Section 3.3, we aim to identify global optimal point using both these solutions.

3.1 Relative queue discipline management parameter $\beta < \infty$

The above analysis establishes that a KKT point should satisfy either condition C1 or condition C2. Also, the constraint (10) is binding at the optimum. In the analysis below, we consider C1 and C2 individually and solve the equality relationship $W_p(\lambda_s, \beta) = S_p$ and Equation (16) for unknown elements of the KKT points. The analysis assuming that KKT point satisfies condition C1 results in Theorem 1 below.

Theorem 1. *Suppose* $\frac{a}{c} > \frac{\lambda_p (2\mu - \lambda_p)}{\mu (\mu - \lambda_p)^2}\psi$. *Then, there exists* $\lambda_s^{(1)}$ *which is the unique root of the cubic* $G(\lambda_s)$ *in the interval* $(0, \mu - \lambda_p)$*:*

$$G(\lambda_s) = 2\mu\lambda_s^3 - [c\psi + \mu(a + 4\phi_0)]\lambda_s^2 + 2\phi_0[c\psi + \mu(a + \phi_0)]\lambda_s - a\mu\phi_0^2 + c\psi\lambda_p(\mu + \phi_0) \quad (25)$$

where $\phi_0 = \mu - \lambda_p$. *Denote* $\lambda_1 = \lambda_p + \lambda_s^{(1)}$ *and further assume that* S_p *lies in the interval* $I \equiv \left[\frac{\psi\lambda_1}{\mu[\mu - \lambda_p]}, \frac{\psi\lambda_1}{[\mu - \lambda_s^{(1)}][\mu - \lambda_1]}\right)$ *and* $\beta^{(1)}$ *is given by*

$$\beta^{(1)} = \begin{cases} \dfrac{[\mu - \lambda_1][\mu S_p[\mu - \lambda_p] - \psi\lambda_1]}{\psi\lambda_1^2 - \mu S_p\lambda_p[\mu - \lambda_1]} & \text{for } \dfrac{\psi\lambda_1}{\mu[\mu - \lambda_p]} \le S_p \le \dfrac{\psi\lambda_1}{\mu[\mu - \lambda_1]}, \\[2ex] \dfrac{S_p\lambda_s^{(1)}[\mu - \lambda_1]}{\psi\lambda_1 - S_p[\mu - \lambda_s^{(1)}][\mu - \lambda_1]} & \text{for } \dfrac{\psi\lambda_1}{\mu[\mu - \lambda_1]} < S_p < \dfrac{\psi\lambda_1}{[\mu - \lambda_s^{(1)}][\mu - \lambda_1]}. \end{cases} \quad (26)$$

Then, $\lambda_s^{(1)}$ *and* $\beta^{(1)}$ *is a strict local maximum of the NLP P1 and the constraint* (10) *is binding at this point.*

Proof. First, we show that $\lambda_s^{(1)}$ is the unique root of the cubic $G(\lambda_s)$ in the interval $(0, \mu - \lambda_p)$.

Claim 1. *If* $\frac{a}{c} > \frac{\lambda_p (2\mu - \lambda_p)}{\mu (\mu - \lambda_p)^2}\psi$, *then, the cubic* $G(\lambda_s)$ *has an unique root in the interval* $(0, \mu - \lambda_p)$.

Proof. See Appendix. □

The work conservation law applied to our setting results in

$$\lambda_s W_s + \lambda_p W_p = \frac{\lambda^2\psi}{\mu[\mu - \lambda]}. \quad (27)$$

Let us assume that the Lagrangian multipliers are $u_1 = -\frac{c\lambda_p}{b}, u_2 = 0$ and $u_3 = 0$. Note that the constraint (10) is binding at the optimum; therefore, these values of the Lagrangian multipliers satisfy KKT conditions (17)-(20). When the Lagrangian multipliers are $u_1 = -\frac{c\lambda_p}{b}$, $u_2 = 0$ and $u_3 = 0$, then the KKT condition (16) can be rewritten as

$$a - 2\lambda_s - c\frac{\partial}{\partial \lambda_s}\left[\lambda_s W_s + \lambda_p W_p\right] = 0. \tag{28}$$

Using equation (27) in equation (28) results in a cubic equation given as

$$G(\lambda_s) \equiv 2\mu\lambda_s^3 - [c\psi + \mu(a + 4\phi_0)]\lambda_s^2 + 2\phi_0[c\psi + \mu(a + \phi_0)]\lambda_s - a\mu\phi_0^2 + c\psi\lambda_p(\mu + \phi_0) = 0$$

where $\phi_0 = \mu - \lambda_p$. As $\lambda_s^{(1)}$ is the unique root of the cubic $G(\lambda_s)$ in the interval $(0, \mu - \lambda_p)$, solving $G(\lambda_s) = 0$ for $\lambda_s \in (0, \mu - \lambda_p)$ results in $\lambda_s = \lambda_s^{(1)}$.

Claim 2. *There exists a queue discipline management parameter $\bar{\beta} \geq 0$ which satisfies the equality $W_p(\lambda_s, \beta) = S_p$ if $\lambda_p \geq 0$, $\lambda_s \geq 0$, $\lambda_p + \lambda_s < \mu$ and S_p lies in the interval $\left[\frac{\psi\lambda}{\mu[\mu - \lambda_p]}, \frac{\psi\lambda}{[\mu - \lambda_s][\mu - \lambda]}\right)$, where $\lambda = \lambda_p + \lambda_s$. The value of $\bar{\beta}$ is*

$$\bar{\beta} = \begin{cases} \dfrac{[\mu - \lambda][\mu S_p[\mu - \lambda_p] - \psi\lambda]}{\psi\lambda^2 - \mu S_p \lambda_p [\mu - \lambda]} & \text{for } \dfrac{\psi\lambda}{\mu[\mu - \lambda_p]} \leq S_p \leq \dfrac{\psi\lambda}{\mu[\mu - \lambda]} \\[2ex] \dfrac{S_p \lambda_s [\mu - \lambda]}{\psi\lambda - S_p[\mu - \lambda_s][\mu - \lambda]} & \text{for } \dfrac{\psi\lambda}{\mu[\mu - \lambda]} < S_p < \dfrac{\psi\lambda}{[\mu - \lambda_s][\mu - \lambda]} \end{cases} \tag{29}$$

Proof. See Appendix. □

Let $\beta^{(1)} = \bar{\beta}$ for $\lambda_s = \lambda_s^{(1)}$ and $S_p \in I$. From Claim 2, we know that $W_p(\lambda_s^{(1)}, \beta^{(1)}) = S_p$. The point given by $\lambda_s^{(1)}$, $\beta^{(1)}$, $u_1 = -\frac{c\lambda_p}{b}$, $u_2 = 0$ and $u_3 = 0$ satisfies KKT conditions (16)-(22) and is a KKT point of the problem P1.

The restricted Lagrangian function $\tilde{L}_1(\lambda_s, \beta)$, (page no. 168, Bazaraa *et al.* (1993)), at this KKT point is given by $L_1\left(\lambda_s, \beta; u_1 = -\frac{c\lambda_p}{b}, u_2 = 0, u_3 = 0\right)$. Using Equations (27) and (15), we get

$$\tilde{L}_1(\lambda_s, \beta) = \frac{1}{b}\left[a\lambda_s - \lambda_s^2 - c\frac{\lambda^2\psi}{\mu[\mu - \lambda]} + c\lambda_p S_p\right]. \tag{30}$$

We note that the restricted Lagrangian function is independent of β. The Hessian of the restricted Lagrangian function $H_{\tilde{L}_1}(\lambda_s, \beta)$ is given by

$$\begin{bmatrix} -\frac{2}{b}\left(1 + \frac{c\mu\psi}{(\mu-\lambda)^3}\right) & 0 \\ 0 & 0 \end{bmatrix}.$$

The above matrix is negative semi-definite as $\mu - \lambda > 0$. It is evident that at $S_p = \frac{\psi\lambda_1}{\mu[\mu - \lambda_p]}$, the constraints $g_1(\lambda_s, \beta) \equiv W_p(\lambda_s, \beta) \leq S_p$ and $g_2(\lambda_s, \beta) \equiv \beta \geq 0$ are binding for this KKT point. Also, the constraint $g_1(\lambda_s, \beta)$ is strongly active (i.e., the associated Lagrangian multiplier is non-zero) whereas the constraint $g_2(\lambda_s, \beta)$ is weakly active (i.e., the associated Lagrangian multiplier is zero). The gradients of these binding constraints are given by

$$\nabla g_1\left(\lambda_s, \beta\right) = \begin{bmatrix} \kappa_1 \\ \\ \kappa_2 \end{bmatrix} \text{ and } \nabla g_2\left(\lambda_s^{(2)}, 0\right) = \begin{bmatrix} 0 \\ \\ 1 \end{bmatrix}$$

where $\kappa_1 = \frac{\partial W_p}{\partial \lambda_s}$ and $\kappa_2 = \frac{\partial W_p}{\partial \beta}$. We know that $\frac{\partial W_p}{\partial \lambda_s}, \frac{\partial W_p}{\partial \beta} > 0$ for $\lambda_s \in (0, \mu - \lambda_p)$ and $\beta \geq 0$. A non-zero vector $d \equiv (d_1, d_2)$ that satisfies $d.\nabla g_1(\lambda_s^{(1)}, \beta^{(1)}) = 0$ and $d.\nabla g_2(\lambda_s^{(1)}, \beta^{(1)}) \geq 0$ simultaneously is given by $d_1 = -\frac{\kappa_2 d_2}{\kappa_1}$ such that $d_2 > 0$. We note that

$$dH_{\tilde{L}_1}\left(\lambda_s^{(1)}, \beta^{(1)}\right) d^T = -\frac{2}{b}\left(1 + \frac{c\mu\psi}{(\mu - \lambda_1)^3}\right)\left(-\frac{\kappa_2 d_2}{\kappa_1}\right)^2 < 0 \qquad \text{for all } \; d_2 > 0.$$

Next, if $S_p > \frac{\psi\lambda_1}{\mu[\mu - \lambda_p]}$, then only constraint $g_1(\lambda_s, \beta) \equiv W_p(\lambda_s, \beta) \leq S_p$ is binding at that KKT point and also it is strongly active. A non-zero vector $d \equiv (d_1, d_2)$ that satisfies $d.\nabla g_1(\lambda_s^{(1)}, \beta^{(1)}) = 0$ is given by $d_1 = -\frac{\kappa_2 d_2}{\kappa_1}$ such that $d_2 \neq 0$. Again, we note that

$$dH_{\tilde{L}_1}\left(\lambda_s^{(1)}, \beta^{(1)}\right) d^T = -\frac{2}{b}\left(1 + \frac{c\mu\psi}{(\mu - \lambda_1)^3}\right)\left(-\frac{\kappa_2 d_2}{\kappa_1}\right)^2 < 0 \qquad \text{for all } \; d_2 \neq 0.$$

Hence, the KKT point $\lambda_s^{(1)}$, $\beta^{(1)}$, $u_1 = -\frac{c\lambda_p}{b}$, $u_2 = 0$ and $u_3 = 0$ is strict local maximum of the NLP P1 if S_p lies in the interval I. □

Corollary 1. *The mean arrival rate of the secondary customer $\lambda_s^{(1)}$ which is a local optima point, is independent of S_p in the interval I.*

Proof. The optimal $\lambda_s^{(1)}$ is the root of cubic $G(\lambda_s)$ which is independent of S_p. Therefore, $\lambda_s^{(1)}$ is independent of S_p in the interval I. □

We now find KKT points which satisfy condition C2. This results in a strict local maximum of the NLP P1 for S_p lying left to the interval I. Theorem 2 below states this result.

Theorem 2. *Suppose $\frac{a}{c} > \frac{\lambda_p\left(2\mu - \lambda_p\right)}{\mu\left(\mu - \lambda_p\right)^2}\psi$ and S_p lies in the interval $I^- \equiv \left(\frac{\psi\lambda_p}{\mu[\mu-\lambda_p]}, \frac{\psi\lambda_1}{\mu[\mu-\lambda_p]}\right)$ where $\lambda_1 = \lambda_p + \lambda_s^{(1)}$ and $\lambda_s^{(1)}$ is the unique root of the cubic $G(\lambda_s)$ of Equation (25) in the interval $(0, \mu - \lambda_p)$. Then, $\lambda_s^{(2)} = \dfrac{\mu[\mu - \lambda_p]S_p}{\psi} - \lambda_p$ and $\beta^{(2)} = 0$ is a strict local maximum of the NLP P1 and the constraint (10) is binding at this point.*

Proof. Let us first assume that the queue discipline management parameter $\beta = 0$ and the Lagrangian multiplier $u_2 = 0$ at the optimum. Note that the constraint (10) is binding at the optimum. Given $\beta = 0$, the equality relationship $W_p(\lambda_s, \beta) = S_p$ results in

$$\lambda_s^{(2)} \equiv \lambda_s = \frac{\mu[\mu - \lambda_p]S_p}{\psi} - \lambda_p. \tag{31}$$

We note that $\lambda_s^{(2)}$ is an increasing function of S_p as $\frac{\partial \lambda_s^{(2)}}{\partial S_p} = \frac{\mu[\mu-\lambda_p]}{\psi} > 0$ for $0 < \lambda_p < \mu$ and $\psi > 0$. Also, $\lambda_s^{(2)} = \lambda_s^{(1)}$ at $S_p = \frac{\psi\lambda_1}{\mu[\mu-\lambda_p]}$. Therefore, $0 < \lambda_s^{(2)} < \lambda_s^{(1)}$ for $S_p \in I^-$. As $u_2 = 0$ and $\beta = 0$, the KKT conditions (16) results in

$$u_1^{(2)} \equiv u_1 = -\left[a - 2\lambda_s^{(2)} - c\psi\frac{\mu(\lambda_2 + \lambda_s^{(2)}) - \lambda_2^2}{(\mu - \lambda_p)(\mu - \lambda_2)^2}\right]\frac{\mu(\mu - \lambda_p)}{b\psi} \tag{32}$$

where $\lambda_2 = \lambda_p + \lambda_s^{(2)}$. Also, we note that if $0 < \lambda_s < \lambda_s^{(1)}$, then

$$\frac{c\lambda_p}{b} - \left[a - 2\lambda_s - c\psi\frac{\mu\left(\lambda + \lambda_s\right) - \lambda^2}{\left(\mu - \lambda_p\right)\left(\mu - \lambda\right)^2}\right]\frac{\mu\left(\mu - \lambda_p\right)}{b\psi} = \frac{(\mu - \lambda_p)G(\lambda_s)}{b\psi\left(\mu - \lambda\right)^2} < 0.$$

This inequality follows from the proof of Claim 1 which establishes that $G(\lambda_s) < 0$ when $0 < \lambda_s < \lambda_s^{(1)}$ and $\frac{a}{c} > \frac{\lambda_p(2\mu-\lambda_p)}{\mu(\mu-\lambda_p)^2}\psi$. This implies that $u_1^{(2)} < -\frac{c\lambda_p}{b}$ as $0 < \lambda_s^{(2)} < \lambda_s^{(1)}$. Take $u_3^{(2)} = u_3$, obtained using $\lambda_s = \lambda_s^{(2)}$ and $u_1 = u_1^{(2)}$ in Equation (24). We note that $u_3^{(2)} > 0$ as $u_1^{(2)} < -\frac{c\lambda_p}{b}$. The point $\lambda_s^{(2)}$, $\beta = 0$, $u_1^{(2)}$, $u_2 = 0$ and $u_3^{(2)}$ satisfies the KKT conditions (16)-(22). Thus, it is a KKT point.

Given $S_p \in I^-$, we note that the constraints $g_1(\lambda_s, \beta) \equiv W_p(\lambda_s, \beta) \leq S_p$ and $g_2(\lambda_s, \beta) \equiv \beta \geq 0$ are binding for this KKT point. Also, these constraints are strongly active. The gradients of these binding constraints at this KKT point are

$$\nabla g_1\left(\lambda_s^{(2)}, 0\right) = \begin{bmatrix} \frac{\psi}{\mu-\lambda_p} \\ \\ \frac{\lambda_s^{(2)}\lambda_2\psi}{(\mu-\lambda_p)(\mu-\lambda_2)^2} \end{bmatrix} \text{ and } \nabla g_2\left(\lambda_s^{(2)}, 0\right) = \begin{bmatrix} 0 \\ \\ 1 \end{bmatrix}$$

where $\lambda_2 = \lambda_p + \lambda_s^{(2)}$. We observe the both terms of $\nabla g_1(\lambda_s^{(2)}, 0)$ are strictly non zero as $\lambda_p, \lambda_s^{(2)} > 0$ and $\lambda_p + \lambda_s^{(2)} < \mu$. Hence, the gradients of these binding constraints, $\nabla g_1(\lambda_s^{(2)}, 0)$ and $\nabla g_2(\lambda_s^{(2)}, 0)$, at the KKT point are linearly independent. Therefore, this KKT point is a strict local maximum (Corollary of Theorem 4.4.2, Bazaraa *et al.* (1993)). □

Corollary 2. *$\lambda_s^{(2)}$ is a linearly increasing function of S_p in the interval I^-.*

Proof. We have $\frac{\partial \lambda_s^{(2)}}{\partial S_p} = \frac{\mu[\mu-\lambda_p]}{\psi} > 0$ as $\mu > \lambda_p > 0$ and $\psi > 0$. Also, $\frac{\partial^2 \lambda_s^{(2)}}{\partial S_p^2} = 0$. □

3.2 Relative queue discipline management parameter $\boldsymbol{\beta = \infty}$

In this section, we analyze the resulting one-dimensional optimization problem by setting $\beta = \infty$ in P1. Let $\tilde{W}_s(\lambda_s) = W_s(\lambda_s, \beta = \infty)$ and $\tilde{W}_p(\lambda_s) = W_p(\lambda_s, \beta = \infty)$. The resulting optimization problem, P2, is given as

$$\textbf{P2}: \qquad \max_{\lambda_s} \ \frac{1}{b}\left[a\lambda_s - \lambda_s^2 - c\lambda_s\tilde{W}_s(\lambda_s)\right] \tag{33}$$

$$\text{Subject to:} \qquad \tilde{W}_p(\lambda_s) \leq S_p \tag{34}$$

$$\lambda_s \leq \mu - \lambda_p \tag{35}$$

$$\lambda_s \geq 0 \tag{36}$$

Let us define $f_1(\lambda_s) = a\lambda_s - \lambda_s^2$ and $f_2(\lambda_s) = \lambda_s\tilde{W}_s(\lambda_s)$. $f_1(\lambda_s)$ is concave function of λ_s as $\frac{\partial^2 f_1}{\partial \lambda_s^2} = -2 < 0$. We observe that $\lambda_s \geq 0$ and $\tilde{W}_s(\lambda_s) \geq 0$ for $\lambda_s \in [0, \mu - \lambda_p)$. We note that λ_s is linear and $\tilde{W}_s(\lambda_s)$ convex increasing functions of λ_s in the interval $[0, \mu - \lambda_p)$. We know that product of two positively valued, increasing convex functions of the real variable defined on the same interval is an increasing convex function. This implies that $f_2(\lambda_s)$ is an increasing convex function of λ_s in the interval $[0, \mu - \lambda_p)$. As objective function is $\frac{1}{b}[f_1(\lambda_s) - cf_2(\lambda_s)]$, it is a concave function of λ_s in the $[0, \mu - \lambda_p)$. Also, we know that $\tilde{W}_p(\lambda_s)$ is a convex function of λ_s in the interval $[0, \mu - \lambda_p)$. So, if the objective function and constraints of the optimization problem P2 satisfy the KKT sufficiency conditions, the KKT point, if it exists, will be a *global* optimum. Further, we note that the first term of the objective function is increasing whereas the last two terms are decreasing functions of λ_s in the interval $[0, \mu)$. Therefore, the objective function is a unimodal function of λ_s in the interval $[0, \mu)$.

Following the earlier arguments, we observe that the constraint (35) will remain *non-binding* at the optimum. The Lagrangian function corresponding to the NLP P2 can be expressed as

$$L_2(\lambda_s, v_1, v_2) = \frac{1}{b}\left[a\lambda_s - \lambda_s^2 - c\lambda_s\tilde{W}_s(\lambda_s)\right] + v_1\left[\tilde{W}_p(\lambda_s) - S_p\right] + v_2\lambda_s \tag{37}$$

where v_1 and v_2 are the Lagrangian multipliers. The KKT first order necessary conditions for the NLP P2 are given as follows:

$$a - 2\lambda_s - c\left[\tilde{W}_s + \lambda_s\frac{\partial \tilde{W}_s}{\partial \lambda_s}\right] + bv_1\frac{\partial \tilde{W}_p}{\partial \lambda_s} + bv_2 = 0 \tag{38}$$

$$v_1\left[\tilde{W}_p - S_p\right] = 0 \tag{39}$$

$$v_2\lambda_s = 0 \tag{40}$$

$$\tilde{W}_p \leq S_p \ \text{ and } \ \lambda_s < \mu - \lambda_p \tag{41}$$

$$v_1 \leq 0; \lambda_s, v_2 \geq 0 \tag{42}$$

A KKT point is defined by a specific (λ_s, v_1, v_2) that satisfies the conditions (38)-(42). Again, note that if the Lagrangian multiplier $v_2 > 0$ then the KKT condition (40) is satisfied if and only if $\lambda_s = 0$, in which case objective function value is zero. As the objective of the resource owner is to earn a strict positive revenue, we ignore values of $v_2 > 0$ in further analysis and assume throughout that $v_2 = 0$. The analysis below look for all possible KKT points of the revenue maximization problem P2 with $v_2 = 0$, i.e., assign specific values to the remaining two unknown elements of a possible KKT point. We also know that the constraint (34) will be either strictly binding or non-binding at the optimum. Theorem 3 identifies an interval of S_p where the constraint (34) is strictly non-binding at optimality.

Theorem 3. *Suppose* $\frac{a}{c} > \frac{\lambda_p}{\mu^2}\psi$ *and* $\frac{\mu - \lambda_p}{\mu\lambda_p} > \frac{a\lambda_p - c\psi}{2\mu\lambda_p^2 + c\psi(\mu + \lambda_p)}$. *Then, there exists* $\lambda_s^{(3)}$ *which is the unique root of the cubic* $\tilde{G}(\lambda_s)$ *in the interval* $(0, \mu - \lambda_p)$:

$$\tilde{G}(\lambda_s) = 2\mu\lambda_s^3 - \left[a\mu + c\psi + 4\mu^2\right]\lambda_s^2 + 2\mu\left[a\mu + c\psi + \mu^2\right]\lambda_s - \mu\left[a\mu^2 - c\psi\lambda_p\right] \qquad (43)$$

Denote $\lambda_3 = \lambda_p + \lambda_s^{(3)}$ *and further assume that* S_p *lies in the interval* $J \equiv \left(\frac{\psi\lambda_3}{[\mu - \lambda_s^{(3)}][\mu - \lambda_3]}, \infty\right)$. *Then,* $\lambda_s^{(3)}$ *is the global maximum of the NLP P2 and the constraint* (34) *is non-binding at this point.*

Proof. First, we show that $\lambda_s^{(3)}$ is the the unique root of the cubic $\tilde{G}(\lambda_s)$ in the interval $(0, \mu)$.

Claim 3. *If* $\frac{a}{c} > \frac{\lambda_p}{\mu^2}\psi$, *then, cubic* $\tilde{G}(\lambda_s)$ *has unique root in in the interval* $(0, \mu)$.

Proof. See Appendix. □

Observe that given $\frac{\mu-\lambda_p}{\mu\lambda_p} > \frac{a\lambda_p - c\psi}{2\mu\lambda_p^2 + c\psi(\mu+\lambda_p)}$, the unique root of cubic $\tilde{G}(\lambda_s)$ in the interval $(0, \mu)$ indeed is strictly less than $\mu - \lambda_p$. This follows from the fact that $\tilde{G}(\mu - \lambda_p) = (\mu - \lambda_p)(2\mu\lambda_p^2 + c\psi(\mu + \lambda_p)) - \mu\lambda_p(a\lambda_p - c\psi) > 0$ as $\frac{\mu-\lambda_p}{\mu\lambda_p} > \frac{a\lambda_p - c\psi}{2\mu\lambda_p^2 + c\psi(\mu+\lambda_p)}$.

Let us assume that the Lagrangian multiplier $v_1 = 0$ at optimum. Given $v_1 = v_2 = 0$, the KKT condition (38) results in a cubic equation given as

$$\tilde{G}(\lambda_s) \equiv 2\mu\lambda_s^3 - \left[a\mu + c\psi + 4\mu^2\right]\lambda_s^2 + 2\mu\left[a\mu + c\psi + \mu^2\right]\lambda_s - \mu\left[a\mu^2 - c\psi\lambda_p\right] = 0$$

Note that under assumptions of $\frac{a}{c} > \frac{\lambda_p}{\mu^2}\psi$ and $\frac{\mu-\lambda_p}{\mu\lambda_p} > \frac{a\lambda_p - c\psi}{2\mu\lambda_p^2 + c\psi(\mu+\lambda_p)}$, $\lambda_s^{(3)}$ is the unique root of the cubic $\tilde{G}(\lambda_s)$ in the interval $(0, \mu - \lambda_p)$. Therefore, solving $\tilde{G}(\lambda_s) = 0$ for $\lambda_s \in (0, \mu - \lambda_p)$ results in $\lambda_s = \lambda_s^{(3)}$. Further, $\tilde{W}_p(\lambda_s^{(3)}) = \frac{\psi\lambda_3}{[\mu - \lambda_s^{(3)}][\mu - \lambda_3]}$ where $\lambda_3 = \lambda_p + \lambda_s^{(3)}$. We note that $\tilde{W}_p(\lambda_s^{(3)}) < S_p$ for $S_p \in J$. The $\lambda_s^{(3)}$, $v_1 = 0$ and $v_2 = 0$ satisfies KKT conditions (38)-(42) and therefore it is a KKT point. This point is global maximum of P2 for $S_p \in J$ as P2 is a convex optimization problem. □

Let us assume that the input parameters satisfy the assumptions of Theorem 3. This implies that $\lambda_s^{(3)} < \mu - \lambda_p$ and interval J is defined. Earlier, we demonstrated that the objective function of the revenue maximization problem P2, say $O_2(\lambda_s)$, is unimodal function of λ_s in the interval $[0, \mu)$. Given $S_p \in J$, each constraint of P2 is strictly non-binding at $\lambda_s = \lambda_s^{(3)}$. Therefore $\lambda_s^{(3)}$ should also correspond to the unimodal point of the objective function. This implies that the objective function is increasing in the interval $[0, \lambda_s^{(3)}]$. Now, we claim that the constraint (34) is binding at optimum for $S_p \notin J$. This is proved using contradiction. Note that $S_p \notin J$ lies left to the interval J. Let us assume that $\bar{\lambda}_s$ is an optimum point for $S_p \notin J$ and the constraint (34) is strictly non-binding at the optimum, i.e., $\tilde{W}_p(\bar{\lambda}_s) < S_p$. Note that, if $\bar{\lambda}_s > \lambda_s^{(3)}$, then $\tilde{W}_p(\lambda_s^{(3)}) < S_p$ also holds as $\tilde{W}_p(\lambda_s)$ is an increasing function of $\lambda_s \in [0, \mu - \lambda_p)$. This contradicts the assumption that $\bar{\lambda}_s$ is optimum because $\lambda_s^{(3)}$ is unimodal point of the objective function. Therefore, $\bar{\lambda}_s < \lambda_s^{(3)}$. Assume that $\lambda_s = \hat{\lambda}_s$ satisfies $\tilde{W}_p(\hat{\lambda}_s) = S_p$ for the same $S_p \notin J$. As $\tilde{W}_p(\lambda_s)$ is an increasing function of $\lambda_s \in [0, \mu - \lambda_p)$, the inequality $\bar{\lambda}_s < \hat{\lambda}_s \leq \lambda_s^{(3)}$ will hold. This results in $O_2(\bar{\lambda}_s) < O_2(\hat{\lambda}_s)$ as the the objective function is increasing in the interval $[0, \lambda_s^{(3)}]$. This contradicts the initial assumption that $\bar{\lambda}_s$ is optimum and the constraint (34) strictly non-binding at the optimum. Hence, the constraint (34) should be binding at optimum for $S_p \notin J$. A graphical illustration of this argument is given in Figure 3.

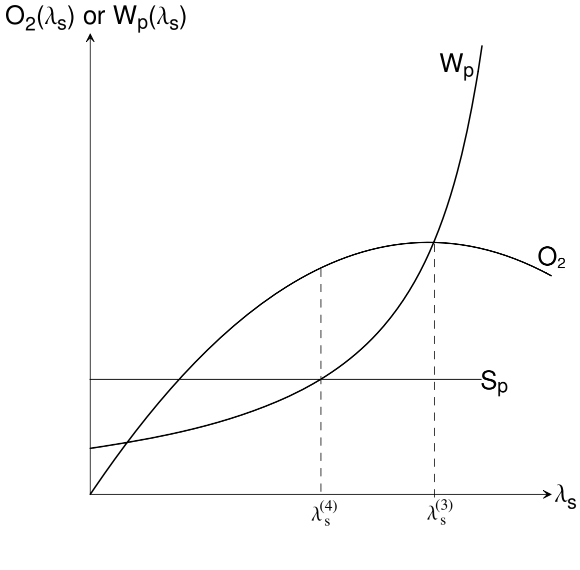


Figure 3: Illustration of the optimization problem P2

Also, we note that if $\frac{\mu-\lambda_p}{\mu\lambda_p} \leq \frac{a\lambda_p - c\psi}{2\mu\lambda_p^2 + c\psi(\mu+\lambda_p)}$, then the unimodal point of the objective function lies outside the interval $[\mu - \lambda_p)$. Again, the similar arguments establish that the constraint (34) should be binding at optimum for $S_p > \hat{S}_p$. The above discussion asserts that the constraint (34) is binding at optimum for $S_p \notin J$. Theorem 4 defines such intervals of S_p where the constraint (34) is binding.

Theorem 4. *Suppose* $\frac{a}{c} > \frac{\lambda_p}{\mu^2}\psi$ *and* S_p *lies in the interval* J^- *that is defined as:*

$$J^- = \begin{cases} \left(\dfrac{\psi\lambda_p}{\mu\left[\mu-\lambda_p\right]}, \dfrac{\psi\lambda_3}{[\mu-\lambda_s^{(3)}][\mu-\lambda_3]}\right] & \textit{if } \dfrac{\mu-\lambda_p}{\mu\lambda_p} > \dfrac{a\lambda_p - c\psi}{2\mu\lambda_p^2 + c\psi(\mu+\lambda_p)}, \\ \left(\dfrac{\psi\lambda_p}{\mu\left[\mu-\lambda_p\right]}, \infty\right) & \textit{otherwise} \end{cases} \tag{44}$$

where $\lambda_3 = \lambda_p + \lambda_s^{(3)}$ *and* $\lambda_s^{(3)}$ *is the unique root of the cubic* $\tilde{G}(\lambda_s)$ *of Equation* (43) *in the interval* $(0, \mu - \lambda_p)$ *whenever* $\frac{\mu-\lambda_p}{\mu\lambda_p} > \frac{a\lambda_p - c\psi}{2\mu\lambda_p^2 + c\psi(\mu+\lambda_p)}$.
Then, $\lambda_s^{(4)} = \frac{1}{2S_p}\left[S_p\left[2\mu-\lambda_p\right] + \psi - \sqrt{\left[S_p\lambda_p + \psi\right]^2 + 4\mu\psi S_p}\right]$ *is the global maximum of the NLP P2 and the constraint* (34) *is binding at this point.*

Proof. We note that $J^- \cap J = \emptyset$; therefore, the constraint (34) will be binding at optimum for $S_p \in J^-$. We first show that there exists unique $\lambda_s \in (0, \mu-\lambda_p)$ that satisfy the equality $\tilde{W}_p(\lambda_s) = S_p$ for $S_p > \frac{\lambda_p\psi}{\mu[\mu-\lambda_p]}$.

Claim 4. *Given* $S_p > \frac{\lambda_p\psi}{\mu[\mu-\lambda_p]}$, *there exists unique* $\tilde{\lambda}_s$ *in the interval* $(0, \mu - \lambda_p)$ *that satisfy the equality* $\tilde{W}_p(\lambda_s) = S_p$. *It is given by*

$$\tilde{\lambda}_s = \frac{1}{2S_p}\left[S_p\left[2\mu-\lambda_p\right] + \psi - \sqrt{\left[S_p\lambda_p + \psi\right]^2 + 4\mu\psi S_p}\right]$$

Proof. See Appendix. □

Take $\lambda_s^{(4)} = \tilde{\lambda}_s$. As $\tilde{W}_p(\lambda_s^{(4)}) = S_p$, the point $\lambda_s = \lambda_s^{(4)}$ satisfies the KKT condition (39) irrespective of the value of the Lagrangian multiplier v_1. Given $\lambda_s = \lambda_s^{(4)}$ and $v_2 = 0$, the KKT condition (38) results in

$$v_1^{(4)} \equiv v_1 = -\left[a - 2\lambda_s^{(4)} - c\psi\frac{\mu\left[\lambda_p + 2\lambda_s^{(4)}\right] - \left[\lambda_s^{(4)}\right]^2}{\mu\left[\mu - \lambda_s^{(4)}\right]^2}\right]\frac{\left[\mu-\lambda_4\right]^2\left[\mu-\lambda_s^{(4)}\right]^2}{b\psi\left[\mu\left(\mu+\lambda_p\right) - \lambda_4^2\right]} \tag{45}$$

where $\lambda_4 = \lambda_p + \lambda_s^{(4)}$. We note that $v_1^{(4)} \leq 0$ if and only if $\frac{a-2\lambda_s^{(4)}}{c\psi} \geq \frac{\mu\left[\lambda_p+2\lambda_s^{(4)}\right]-\left[\lambda_s^{(4)}\right]^2}{\mu\left[\mu-\lambda_s^{(4)}\right]^2}$ as $\lambda_p > 0$ and $0 < \lambda_s^{(4)} < \mu - \lambda_p$. The rearrangement of this inequality results in $\tilde{G}(\lambda_s^{(4)}) \leq 0$. Given $\frac{a}{c} > \frac{\lambda_p}{\mu^2}\psi$, the Claim 3 has already established that $\tilde{G}(\lambda_s) \leq 0$ in the interval $(0, \lambda_s^{(3)}]$ where $\lambda_s^{(3)}$ is the unique root of the cubic $\tilde{G}(\lambda_s)$ in the interval $(0, \mu)$. This implies that if $\lambda_s^{(4)} \leq \lambda_s^{(3)}$ then the inequality $v_1^{(4)} \leq 0$ will hold true.

To establish that $\lambda_s^{(4)} \leq \lambda_s^{(3)}$, first consider the case when $\frac{\mu-\lambda_p}{\mu\lambda_p} \leq \frac{a\lambda_p - c\psi}{2\mu\lambda_p^2 + c\psi(\mu+\lambda_p)}$ and thereby the interval $J^- = \left(\frac{\lambda_p\psi}{\mu[\mu-\lambda_p]}, \infty\right)$. Given $\frac{\mu-\lambda_p}{\mu\lambda_p} \leq \frac{a\lambda_p - c\psi}{2\mu\lambda_p^2 + c\psi(\mu+\lambda_p)}$, we note from the proof of Theorem 3 that $\lambda_s^{(3)}$, the root of the cubic $\tilde{G}(\lambda_s)$, will always be greater than or equal to $\mu - \lambda_p$. We have established in the Claim 4 that $\lambda_s^{(4)} < \mu - \lambda_p$ and therefore $\lambda_s^{(4)} < \lambda_s^{(3)}$ under this assumption. Next, consider the case when $\frac{\mu-\lambda_p}{\mu\lambda_p} > \frac{a\lambda_p - c\psi}{2\mu\lambda_p^2 + c\psi(\mu+\lambda_p)}$ and thereby the interval $J^- = \left(\frac{\lambda_p\psi}{\mu[\mu-\lambda_p]}, \frac{\psi\tilde{\lambda}}{[\mu-\lambda_s^{(3)}][\mu-\tilde{\lambda}]}\right]$. As the constraint (34) is binding at optimum, we note that $\lambda_s^{(4)} = \lambda_s^{(3)}$ at $S_p = \frac{\psi\tilde{\lambda}}{[\mu-\lambda_s^{(3)}][\mu-\tilde{\lambda}]}$. We also know that $\tilde{W}_p(\lambda_s)$ is an increasing convex function of λ_s in the interval $[0, \mu-\lambda_p)$. Therefore, $S_p < \frac{\psi\tilde{\lambda}}{[\mu-\lambda_s^{(3)}][\mu-\tilde{\lambda}]}$ will always result in $\lambda_s^{(4)} < \lambda_s^{(3)}$. This completes the argument that $\lambda_s^{(4)} \leq \lambda_s^{(3)}$ for $S_p \in J^-$. The $\lambda_s = \lambda_s^{(4)}$, $v_1 = v_1^{(4)}$, $v_2 = 0$ satisfies KKT conditions (38)-(42). Therefore, it is a KKT point and thereby a global maximum of the optimization problem P2. □

Corollary 3. *$\lambda_s^{(4)}$ is an increasing function of S_p for $S_p \in J^-$.*

Proof. Given $\lambda_s^{(4)}$, as defined in Theorem 4, we note that

$$\frac{\partial\lambda_s^{(4)}}{\partial S_p} = -\frac{\psi}{2S_p^2}\left[1 - \frac{S_p(2\mu+\lambda_p)+\psi}{\sqrt{[S_p\lambda_p+\psi]^2 + 4\psi\mu S_p}}\right].$$

We note that $\frac{\partial\lambda_s^{(4)}}{\partial S_p} \geq 0$ as

$$\begin{aligned}[S_p(2\mu+\lambda_p)+\psi]^2 &= [S_p\lambda_p+\psi]^2 + 4\psi\mu S_p + 4\mu(\mu+\lambda_p)S_p^2 \\ &\geq [S_p\lambda_p+\psi]^2 + 4\psi\mu S_p\end{aligned}$$

which implies that $\left|\frac{S_p(2\mu+\lambda_p)+\psi}{\sqrt{[S_p\lambda_p+\psi]^2+4\psi\mu S_p}}\right| \geq 1$. □

3.3 Search for global optima

The analysis in Section 3.1 establishes that if $\frac{a}{c} > \frac{\lambda_p(2\mu-\lambda_p)}{\mu(\mu-\lambda_p)^2}\psi$, then the optimization problem P1 will have a local optimal solution with $\beta^* < \infty$ provided that $S_p \in I^- \cup I$. The analysis in Section 3.2 establishes that if $\frac{a}{c} > \frac{\lambda_p}{\mu^2}\psi$, then the optimization problem P2 will have a local optimal solution for $S_p > \hat{S}_p = \frac{\lambda_p\psi}{\mu[\mu-\lambda_p]}$. Note that the local optimal solution of P2 also corresponds to the local optimal solution of the optimization problem P1 with $\beta^* = \infty$. Further, we observe that

$$\frac{a}{c} > \frac{\lambda_p\,(2\mu-\lambda_p)}{\mu\,(\mu-\lambda_p)^2}\psi = \left[\frac{2}{\mu-\lambda_p} + \frac{\lambda_p}{(\mu-\lambda_p)^2}\right]\frac{\lambda_p}{\mu}\psi > \frac{\lambda_p}{\mu(\mu-\lambda_p)}\psi > \frac{\lambda_p}{\mu^2}\psi.$$

The above inequalities follow as $0 < \lambda_p < \mu$. The above inequality implies that if $\frac{a}{c} > \frac{\lambda_p(2\mu-\lambda_p)}{\mu(\mu-\lambda_p)^2}\psi$ then $\frac{a}{c} > \frac{\lambda_p}{\mu^2}\psi$ automatically holds. Also, if $\frac{a}{c} \leq \frac{\lambda_p}{\mu^2}\psi$, then the roots of cubics $G(\lambda_s)$ and $\tilde{G}(\lambda_s)$ are negative and doesn't constitute feasible points for the optimization problems P1 and P2. Given $S_p > \hat{S}_p$, the relationship among the input parameters results in the following possibilities:

D1: $\frac{a}{c} \leq \frac{\lambda_p}{\mu^2}\psi$: There does not exist optimum solutions to the optimization problems P1 and P2.

D2: $\frac{\lambda_p}{\mu^2}\psi < \frac{a}{c} \leq \frac{\lambda_p(2\mu-\lambda_p)}{\mu(\mu-\lambda_p)^2}\psi$: There exists an optimum solution to the optimization problem P2, but there does exist any optimum solution to the optimization problem P1.

D3: $\frac{a}{c} > \frac{\lambda_p(2\mu-\lambda_p)}{\mu(\mu-\lambda_p)^2}\psi$: There exist optimum solutions to both the optimization problems P1 and P2 for $S_p \in I^- \cup I$. Equivalently, the original optimization problem P1 has two local optimal solutions; one with $\beta^* < \infty$ and another with $\beta^* = \infty$. But, for $S_p > I_u$ where $I_u = \frac{\psi\lambda_1}{[\mu-\lambda_s^{(1)}][\mu-\lambda_1]}$ is the upper limit of the interval I, there exists an optimal solution to the optimization problem P2 only.

Given $\frac{a}{c} > \frac{\lambda_p(2\mu-\lambda_p)}{\mu(\mu-\lambda_p)^2}\psi$, let $\left(\lambda_s^{\mathsf{f}}, \beta^{\mathsf{f}}\right)$ and $\left(\lambda_s^{\mathsf{i}}, \infty\right)$ are optimal solutions of the optimization problems P1 and P2 respectively for given $S_p \in I^- \cup I$. Also, let the corresponding values of objective function are $O_1^*\left(\lambda_s^{\mathsf{f}}, \beta^{\mathsf{f}}\right)$ and $O_2^*\left(\lambda_s^{\mathsf{i}}, \infty\right)$. Below, we *seek* to establish that $O_1^*\left(\lambda_s^{\mathsf{f}}, \beta^{\mathsf{f}}\right) \geq O_2^*\left(\lambda_s^{\mathsf{i}}, \infty\right)$.

First, observe from Theorems 3-4 that the feasible region of S_p, i.e., $(\hat{S}_p, \infty)$, is divided into intervals J^- and J. Note that each of the constraints of the optimization problem P2 is non-binding at optimum for $S_p \in J$, while, constraint (34) of the optimization problem P2 is binding at optimum for $S_p \in J^-$. Further, if $\frac{\mu-\lambda_p}{\mu\lambda_p} > \frac{a\lambda_p - c\psi}{2\mu\lambda_p^2 + c\psi(\mu+\lambda_p)}$, then intervals $J^- = (\hat{S}_p, J_\ell]$ and $J = (J_\ell, \infty)$ where $J_\ell = \frac{\psi\lambda_3}{[\mu-\lambda_s^{(3)}][\mu-\lambda_3]}$. On the other hand, if $\frac{\mu-\lambda_p}{\mu\lambda_p} \leq \frac{a\lambda_p - c\psi}{2\mu\lambda_p^2 + c\psi(\mu+\lambda_p)}$, then $J^- = (\hat{S}_p, \infty)$ and $J = \emptyset$. Given $\frac{\mu-\lambda_p}{\mu\lambda_p} > \frac{a\lambda_p - c\psi}{2\mu\lambda_p^2 + c\psi(\mu+\lambda_p)}$, we will now establish that $J_\ell > I_u$, i.e., J_ℓ lies to the right of I_u. This means that for $S_p \in I^- \cup I$, in both the local solutions given by optimization problems P1 and P2 the service level constraint corresponding to primary class customers is binding. We note that $I_u = \xi(\lambda_s^{(1)})$ and $J_\ell = \xi(\lambda_s^{(3)})$ where $\xi(\lambda_s) = \frac{\psi\lambda}{[\mu-\lambda_s][\mu-\lambda]}$ and $\lambda = \lambda_p + \lambda_s$. As $\frac{\partial\xi(\lambda_s)}{\partial\lambda_s} = \frac{[\mu(\mu+\lambda_p)-\lambda^2]\psi}{[\mu-\lambda_s]^2[\mu-\lambda]^2} > 0$ for $\lambda_p > 0$, $\lambda_s > 0$ and $\lambda_p + \lambda_s < \mu$, the inequality $J_\ell > I_u$ will hold if $\lambda_s^{(3)} > \lambda_s^{(1)}$. We argue below that $\lambda_s^{(3)} > \lambda_s^{(1)}$.

We observe from Claim 1 that $\lambda_s^{(1)}$ is the unique root of the cubic $G(\lambda_s)$ of Equation (25) in the interval $(0, \mu - \lambda_p)$ whenever $\frac{a}{c} > \frac{\lambda_p(2\mu-\lambda_p)}{\mu(\mu-\lambda_p)^2}\psi$. That is, $\lambda_s^{(1)} \in (0, \mu - \lambda_p)$ for $a \in (a_\ell, \infty)$ where $a_\ell = \frac{\lambda_p(2\mu-\lambda_p)}{\mu(\mu-\lambda_p)^2}c\psi$. We further observe from Claim 3 and proof of Theorem 3 that $\lambda_s^{(3)}$ is the unique root of the cubic $\tilde{G}(\lambda_s)$ of Equation (43) in the interval $(0, \mu - \lambda_p)$ whenever $\frac{\mu-\lambda_p}{\mu\lambda_p} > \frac{a\lambda_p - c\psi}{2\mu\lambda_p^2 + c\psi(\mu+\lambda_p)}$ and $\frac{a}{c} > \frac{\lambda_p}{\mu^2}\psi$. That is, $\lambda_s^{(3)} \in (0, \mu - \lambda_p)$ for $a \in (\tilde{a}_\ell, \tilde{a}_u)$ where $\tilde{a}_\ell = \frac{\lambda_p}{\mu^2}c\psi$ and $\tilde{a}_u = 2(\mu - \lambda_p) + \frac{c\psi}{\lambda_p}\left[1 + \frac{(\mu-\lambda_p)^2}{\mu\lambda_p}\right]$. We also note that if $\frac{\mu-\lambda_p}{\mu\lambda_p} \leq \frac{a\lambda_p - c\psi}{2\mu\lambda_p^2 + c\psi(\mu+\lambda_p)}$, i.e., $a \geq \tilde{a}_u$, then $\mu - \lambda_p \leq \lambda_s^{(3)} < \mu$. From $\frac{\lambda_p}{\mu^2}\psi < \frac{\lambda_p(2\mu-\lambda_p)}{\mu(\mu-\lambda_p)^2}\psi$, we have that $\tilde{a}_\ell < a_\ell$. Note that $\lambda_s^{(1)} = 0$ at $a = a_\ell$, $\lambda_s^{(3)} = 0$ at $a = \tilde{a}_\ell$ and $\lambda_s^{(3)} = \mu - \lambda_p$ at $a = \tilde{a}_u$. Based on relative values of a_ℓ, $\tilde{a}_\ell$ and $\tilde{a}_u$, we observe following:

- If $a_\ell < \tilde{a}_u$, then

  1. $\lambda_s^{(1)} \leq 0$ and $0 < \lambda_s^{(3)} < \mu - \lambda_p$ for $a \in (\tilde{a}_\ell, a_\ell]$.
  2. $0 < \lambda_s^{(1)} < \mu - \lambda_p$ and $0 < \lambda_s^{(3)} < \mu - \lambda_p$ for $a \in (a_\ell, \tilde{a}_u)$.
  3. $0 < \lambda_s^{(1)} < \mu - \lambda_p$ and $\mu - \lambda_p \leq \lambda_s^{(3)} < \mu$ for $a \geq \tilde{a}_u$.

- If $a_\ell \geq \tilde{a}_u$, then

  1. $\lambda_s^{(1)} < 0$ and $0 < \lambda_s^{(3)} < \mu - \lambda_p$ for $a \in (\tilde{a}_\ell, \tilde{a}_u)$.
  2. $\lambda_s^{(1)} \leq 0$ and $\mu - \lambda_p \leq \lambda_s^{(3)} < \mu$ for $a \in [\tilde{a}_u, a_\ell]$.
  3. $0 < \lambda_s^{(1)} < \mu - \lambda_p$ and $\mu - \lambda_p < \lambda_s^{(3)} < \mu$ for $a > a_\ell$.

It is evident from above that $\lambda_s^{(1)} < \lambda_s^{(3)}$ in all possible values of a except when $a_\ell < \tilde{a}_u$ and $a \in (a_\ell, \tilde{a}_u)$. When $a \in (a_\ell, \tilde{a}_u)$ and $a_\ell < \tilde{a}_u$, then both $\lambda_s^{(1)}$ and $\lambda_s^{(3)}$ lie in $(0, \mu - \lambda_p)$. Given $a_\ell < \tilde{a}_u$, we notice that $\lambda_s^{(3)} > \lambda_s^{(1)}$ at $a = \tilde{a}_u$ as $\lambda_s^{(1)} < \mu - \lambda_p$ and $\lambda_s^{(3)} = \mu - \lambda_p$. Below, we will establish that $\lambda_s^{(3)} < \lambda_s^{(1)}$ for $a \in (a_\ell, \tilde{a}_u)$ whenever $a_\ell < \tilde{a}_u$.

Claim 5. *The root of the cubic $G(\lambda_s)$, $\lambda_s^{(1)}$, is an increasing function of demand function coefficient a.*

Proof. See Appendix. □

Similar arguments result that the root of the cubic $\tilde{G}(\lambda_s)$, $\lambda_s^{(3)}$, is increasing function of a. Note that $\frac{\partial \tilde{G}(\lambda_s, a)}{\partial a} = -\mu(\mu - \lambda_s)^2 < 0$. The Figure 4 illustrates variations in $\lambda_s^{(1)}$

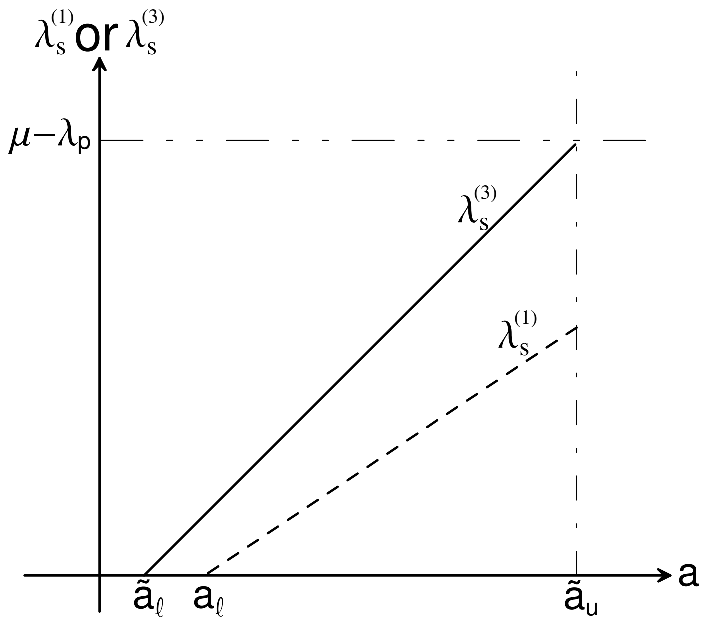


Figure 4: Variation in $\lambda_s^{(1)}$ and $\lambda_s^{(3)}$ with respect to a

and $\lambda_s^{(3)}$ with respect to a in $a \in (\tilde{a}_\ell, \tilde{a}_u)$. As $\frac{\partial G(\lambda_s, a)}{\partial a}$ and $\frac{\partial \tilde{G}(\lambda_s, a)}{\partial a}$ is independent of a, the gap between $\lambda_s^{(3)}$ and $\lambda_s^{(1)}$ will always exist. This implies that $\lambda_s^{(3)} > \lambda_s^{(1)}$. The possible orderings of a_ℓ, $\tilde{a}_\ell$ and $\tilde{a}_u$ determine the existence of feasible $\lambda_s^{(1)}$ and $\lambda_s^{(3)}$ for a given input parameter a. Note that the intervals of S_p, as defined in Theorems 1-4, depend on these values of $\lambda_s^{(1)}$ and $\lambda_s^{(3)}$ and the optimum choice of the decision variables depend on these intervals of S_p. We consider all possible cases in Algorithm.

Suppose $\frac{a}{c} > \frac{\lambda_p(2\mu-\lambda_p)}{\mu(\mu-\lambda_p)^2}\psi$ and let $\left(\lambda_s^{\mathsf{f}}, \beta^{\mathsf{f}}\right)$ and $\left(\lambda_s^{\mathsf{i}}, \infty\right)$ are optimal solutions of the optimization problems P1 and P2 respectively for given $S_p \in I^- \cup I$. Also, let the corresponding values of objective function are $O_1^*\left(\lambda_s^{\mathsf{f}}, \beta^{\mathsf{f}}\right)$ and $O_2^*\left(\lambda_s^{\mathsf{i}}, \infty\right)$. Above we established that $J^\ell > I_u$. This implies that optimization problem P2 has constraint (34) binding at optimum for $S_p \in I^- \cup I$. From the interpretation of the Lagrangian duality, it is known that the marginal rate of change in the objective function value due to incremental increase in the right hand side coefficient of the constraint is given by negative of the Lagrangian multiplier value at the optimality, provided that the KKT point is a regular point (i.e., gradients of the binding constraints are linearly independent). We note from the proofs of Theorems 1 and 2 that the local optimum points corresponding to the optimization problem P1 are regular points. Also, the optimization problem P2 has only one binding constraint at the global optimum point. Thus,

$$\frac{\partial O_1^*}{\partial S_p} = -u_1^{\mathsf{f}} \quad \text{and} \quad \frac{\partial O_2^*}{\partial S_p} = -v_1^{\mathsf{i}}$$

where u_1^{f} and v_1^{i} are the corresponding values of the Lagrangian multipliers associated with the constraint $W_p(\lambda_s, \beta) = S_p$ of the optimization problems P1 and P2 respectively. The rearrangement of Equations (32) and (45) result in

$$u_1^{\mathsf{f}} = \frac{(\mu-\lambda_p)\, G(\lambda_s^{\mathsf{f}})}{b\psi\,(\mu-\lambda_p-\lambda_s^{\mathsf{f}})^2} - \frac{c\lambda_p}{b} \qquad \text{and} \qquad v_1^{\mathsf{i}} = \frac{\left(\mu-\lambda_p-\lambda_s^{\mathsf{i}}\right)^2 \tilde{G}(\lambda_s^{\mathsf{i}})}{b\psi\mu\left[\mu\,(\mu+\lambda_p) - (\lambda_p+\lambda_s^{\mathsf{i}})^2\right]}.$$

We note that $\lambda_s^{\mathsf{f}} = \lambda_s^{(1)}$ for $S_p \in I$ where $\lambda_s^{(1)}$ is the root of the cubic $G(\lambda_s)$. Therefore, $u_1^{\mathsf{f}} = -\frac{c\lambda_p}{b}$ in the interval I. As $u_1^{\mathsf{f}}, v_1^{\mathsf{i}} \le 0$, it implies that both $O_1^*\left(\lambda_s^{\mathsf{f}}, \beta^{\mathsf{f}}\right)$ and $O_2^*\left(\lambda_s^{\mathsf{i}}, \infty\right)$ are increasing functions of S_p. Also,

$$\frac{\partial u_1^{\mathsf{f}}}{\partial \lambda_s^{\mathsf{f}}} = \frac{2\mu\,(\mu-\lambda_p)}{b\psi}\left[1 + \frac{c\mu\psi}{(\mu-\lambda_p-\lambda_s^{\mathsf{f}})}\right] \ge 0$$

$$\frac{\partial v_1^{\mathsf{i}}}{\partial \lambda_s^{\mathsf{i}}} = \frac{1}{b\psi\mu}\left[-\frac{2\mu\left(\mu-\lambda_s^{\mathsf{i}}\right)\left(\mu-\lambda_p-\lambda_s^{\mathsf{i}}\right)}{\left[\mu\,(\mu+\lambda_p)-(\lambda_p+\lambda_s^{\mathsf{i}})^2\right]^2}\tilde{G}(\lambda_s^{\mathsf{i}}) + \frac{\left(\mu-\lambda_p-\lambda_s^{\mathsf{i}}\right)^2}{\left[\mu\,(\mu+\lambda_p)-(\lambda_p+\lambda_s^{\mathsf{i}})^2\right]}\tilde{G}'(\lambda_s^{\mathsf{i}})\right] \ge 0.$$

The above inequalities follow as $\lambda_s^{\mathsf{f}}, \lambda_s^{\mathsf{i}} < \mu - \lambda_p$, $\lambda_s^{\mathsf{i}} \le \lambda_s^{(3)}$ where $\lambda_s^{(3)}$ is the root of the cubic $\tilde{G}(\lambda_s)$. Also, $\tilde{G}(\lambda_s) \le 0$ and is an increasing function of λ_s for $0 \le \lambda_s \le \lambda_s^{(3)}$.

Further,

$$\frac{\partial^2 O_1^*}{\partial S_p^2} = -\frac{\partial u_1^{\mathsf{f}}}{\partial S_p} = -\frac{\partial u_1^{\mathsf{f}}}{\partial \lambda_s^{\mathsf{f}}}\frac{\partial \lambda_s^{\mathsf{f}}}{\partial S_p} \quad \text{and} \quad \frac{\partial^2 O_2^*}{\partial S_p^2} = -\frac{\partial v_1^{\mathsf{i}}}{\partial S_p} = -\frac{\partial v_1^{\mathsf{i}}}{\partial \lambda_s^{\mathsf{i}}}\frac{\partial \lambda_s^{\mathsf{i}}}{\partial S_p}.$$

We note from the results of the Corollaries (1) and (2) that $\frac{\partial \lambda_s^{\mathsf{f}}}{\partial S_p} > 0$ in the interval I^- and $\frac{\partial \lambda_s^{\mathsf{f}}}{\partial S_p} = 0$ in the interval I. Therefore, $\frac{\partial^2 O_1^*}{\partial S_p^2} = -\frac{\partial u_1^{\mathsf{f}}}{\partial S_p} < 0$ and $\frac{\partial^2 O_1^*}{\partial S_p^2} = -\frac{\partial u_1^{\mathsf{f}}}{\partial S_p} = 0$ in the intervals I^- and I respectively. This implies that O_1^* is an increasing concave function of S_p in the interval I^- and is a linearly increasing function of S_p in the interval I. Also, the slope of O_1^* with respect to S_p is decreasing in the interval I^- and remains constant in the interval I. Similarly, we note from the result of the Corollary (3) that $\frac{\partial \lambda_s^{\mathsf{i}}}{\partial S_p} \geq 0$. Therefore, $\frac{\partial^2 O_2^*}{\partial S_p^2} = -\frac{\partial v_1^{\mathsf{i}}}{\partial S_p} \leq 0$. This implies that O_2^* is an increasing concave function of S_p and the slope of O_2^* is decreasing function of S_p.

Note that $\beta^{\mathsf{f}} \to \infty$ as $S_p \to I_u$, the upper limit of the interval I. Given $S_p > \hat{S}_p$, the equality $W_p(\lambda_s, \infty) = S_p$ results in a quadratic equation of λ_s with unique root in the interval $(0, \mu - \lambda_p)$. This implies that $\lambda_s^{\mathsf{f}} \to \lambda_s^{\mathsf{i}}$ and thereby $O_1^*\left(\lambda_s^{\mathsf{f}}, \beta^{\mathsf{f}}\right) \to O_2^*\left(\lambda_s^{\mathsf{i}}, \infty\right)$ as $S_p \to I_u$. Both optimization problems P1 and P2 are identical at $S_p = I_u$ and therefore $v_1^{\mathsf{i}} \to u_1^{\mathsf{f}}$, i.e., $\frac{\partial O_1^*}{\partial S_p} \to \frac{\partial O_2^*}{\partial S_p}$ as $S_p \to I_u$. As slope of $O_2^*\left(\lambda_s^{\mathsf{i}}, \infty\right)$ is a decreasing function of S_p whereas slope of $O_1^*\left(\lambda_s^{\mathsf{f}}, \beta^{\mathsf{f}}\right)$ remains constant in the interval I. Therefore, $O_2^*\left(\lambda_s^{\mathsf{i}}, \infty\right) < O_1^*\left(\lambda_s^{\mathsf{f}}, \beta^{\mathsf{f}}\right)$ in the interval I as these curves intersect at point I_u.

Further, at $S_p = \hat{S}_p$, $\lambda_s^{\mathsf{f}} = \lambda_s^{\mathsf{i}} = 0$, $O_1^* = O_2^* = 0$ and

$$\begin{aligned}
\left.\frac{\partial O_1^*}{\partial S_p}\right|_{\hat{S}_p} - \left.\frac{\partial O_2^*}{\partial S_p}\right|_{\hat{S}_p} &= -u_1^{\mathsf{f}}\Big|_{\lambda_s^{\mathsf{f}}=0} + v_1^{\mathsf{i}}\Big|_{\lambda_s^{\mathsf{i}}=0} \\
&= \frac{1}{b\psi}\left[\frac{\mu\left[a\left(\mu-\lambda_p\right)^2 - c\psi\lambda_p\right]}{\mu-\lambda_p} + \frac{\left(\mu-\lambda_p\right)^2\left[-a\mu^2 + c\psi\lambda_p\right]}{\mu\left(\mu+\lambda_p\right) - \lambda_p^2}\right] \\
&= \frac{1}{b\psi}\left[\frac{\lambda_p(2\mu-\lambda_p)\left[a\mu\left(\mu-\lambda_p\right)^2 - c\psi\lambda_p(2\mu-\lambda_p)\right]}{\left(\mu-\lambda_p\right)\left[\mu\left(\mu+\lambda_p\right)-\lambda_p^2\right]}\right] > 0.
\end{aligned}$$

The last inequality follows as $\lambda_p < \mu$ and $\frac{a}{c} > \frac{\lambda_p(2\mu-\lambda_p)}{\mu(\mu-\lambda_p)^2}\psi$. Therefore, $O_1^* > O_2^*$ at $\hat{S}_p + \epsilon$ where ϵ is a small positive number. Figure 5 illustrates objective function values of P1 and P2 at optimum in the interval $I^- \cup I$ given that $\frac{a}{c} > \frac{\lambda_p(2\mu-\lambda_p)}{\mu(\mu-\lambda_p)^2}\psi$. The above analysis only suggests that the $O_2^*\left(\lambda_s^{\mathsf{i}}, \infty\right) < O_1^*\left(\lambda_s^{\mathsf{f}}, \beta^{\mathsf{f}}\right)$ at $\hat{S}_p + \epsilon$ where ϵ is a small positive number. We summarise these conclusions below.

Theorem 5. *1. Suppose $\frac{a}{c} \leq \frac{\lambda_p}{\mu^2}\psi$. Then, the constrained resource sharing optimization problem P0 is infeasible for $S_p \in (\hat{S}_p, \infty)$.*

2. Suppose $\frac{\lambda_p}{\mu^2}\psi < \frac{a}{c} \leq \frac{\lambda_p(2\mu-\lambda_p)}{\mu(\mu-\lambda_p)^2}\psi$. Then, we can write $(\hat{S}_p, \infty)$ as $(\hat{S}_p, \infty) = J^- \cup J$ with interval J being possibly empty. Then, optimization problem P2 has a solution but optimization problem P1 is infeasible. For $S_p \in (\hat{S}_p, \infty)$, the optimal solution to

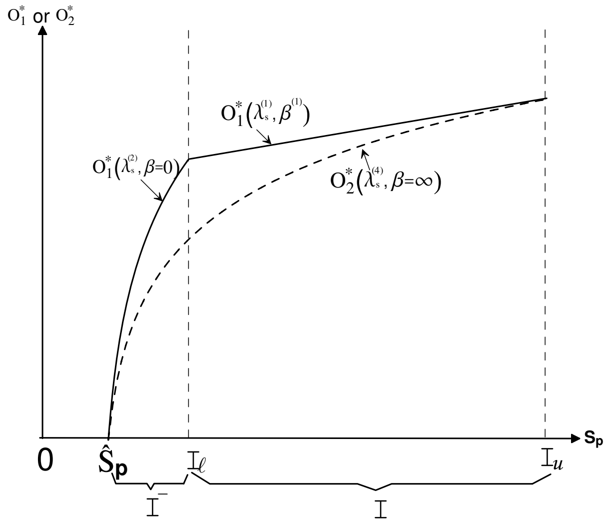


Figure 5: Optimum values of P1 and P2 in interval $I^- \cup I$ whenever $\frac{a}{c} > \frac{\lambda_p(2\mu-\lambda_p)}{\mu(\mu-\lambda_p)^2}\psi$. J_ℓ (possibly infinity) always lies right to I_u.

P0 is given by optimal solutions to P2 with $\beta^* = \infty$ and λ_s^* is either $\lambda_s^{(3)}$ or $\lambda_s^{(4)}$.

3. *Suppose $\frac{a}{c} > \frac{\lambda_p(2\mu-\lambda_p)}{\mu(\mu-\lambda_p)^2}\psi$. Then, we can write $(\hat{S}_p, \infty)$ as $(\hat{S}_p, \infty) = I^- \cup I \cup I^+ \cup J$ with possibly J being an empty interval. Then, optimization problem P1 and P2 have optimal solutions. There exists an $\epsilon > 0$ such that for $S_p \in (\hat{S}_p, \hat{S}_p + \epsilon) \cup I$ the optimal solutions to P0 is given by optimal solution to P1 with $\beta^* < \infty$ and λ_s^* is either $\lambda_s^{(1)}$ or $\lambda_s^{(2)}$. For $S_p \in I^+ \cup J$, the optimal solution to P0 is given by the optimal solution to P2 with $\beta^* = \infty$ and λ_s^* is either $\lambda_s^{(3)}$ or $\lambda_s^{(4)}$.*

It is possible that the $O_2^*\left(\lambda_s^{\mathsf{i}}, \infty\right) > O_1^*\left(\lambda_s^{\mathsf{f}}, \beta^{\mathsf{f}}\right)$ in the interval I^-. We are not able to verify analytically that the above inequality will never hold, but our numerical experiments suggest that $O_2^*\left(\lambda_s^{\mathsf{i}}, \infty\right) < O_1^*\left(\lambda_s^{\mathsf{f}}, \beta^{\mathsf{f}}\right)$ in the interval I^- always. Based on this, we have the following conjecture and remark.

Conjecture. *For $S_p \in I^-$, the optimal solution of P0 is given by optimal solution of P1*

Remark. *We assume henceforth in arriving at an algorithm and in our computations that the conjecture is true.*

4 Algorithm and numerical illustrations

Based on the earlier analysis, we propose an algorithm, that converges in finite steps, for selection of the optimum mean arrival rate of secondary class customers $\lambda_s^* > 0$ and the relative priority queue discipline management parameter β^* that maximizes the revenue of the resource owner while ensuring an agreed upon service level to the primary class customers. We recall that both contract parameters, the optimum assured service level to the secondary class customers S_s^* and the optimal unit admission price charged to the secondary class customers θ^* can be computed as $S_s^* = W_s(\lambda_s^*, \beta^*)$ and $\theta^* = [a - cS_s^* - \lambda_s^*]/b$. The algorithm finds the contract parameters by finding the optimal points of non-convex optimization problem P0 in closed form expressions.

4.1 Algorithm

We have noted in Section 2 that even a dedicated resource will be unable to meet the prevailing service level commitment S_p to the primary class customers if $S_p < \hat{S}_p$ whereas just able to meet this for $S_p = \hat{S}_p$. Therefore, the inclusion of the secondary class customers into the system is possible if and only if $S_p > \hat{S}_p$. This condition corresponds to the system capability. We also noted that $\lambda_s \geq 0$ if and only if $S_s \leq \hat{S}_s(\theta)$ where $\hat{S}_s(\theta) = \frac{a-b\theta}{c}$. Therefore even a free service to secondary class customers, i.e., $\theta = 0$, will be unable to result in $\lambda_s > 0$ for $S_s \geq \frac{a}{c}$. The best value of the assured service level to secondary class customers is achieved by $\beta = \infty$, i.e., assigning the static high priority to the secondary class customers. This follows from the fact that $W_s(\lambda_s, \beta)$ is a decreasing function of β. We have $W_s(\lambda_s = \epsilon, \infty) \approx \frac{\lambda_p}{\mu^2}\psi$ where ϵ is strictly positive and $\epsilon \approx 0$. This implies that $\lambda_s > 0$ if and only if $\frac{a}{c} > \frac{\lambda_p}{\mu^2}\psi$. This condition captures economic viability of secondary class customers. A feasible solution of the revenue maximization problem is possible if and only if both system capability and economic viability is satisfied for a given set of input parameters, i.e. λ_p, μ, σ, a, b, c and S_p. Step one of the algorithm demonstrates this. This also follows from the first point of Theorem 5.

Step 2 of the algorithm describes the possibility of having a unconstrained solution of the optimization problem P0. This follows from Theorem 3. Note that each constraint of the optimization problem P0 remains strictly non-binding at the optimum point described in Theorem 3. Step 3 and Steps 4-5 of the algorithm follows from second and third parts of Theorem 5 respectively. Note that the Step 5(a) follows from the conjecture. The algorithm is given as follows:

Inputs: λ_p, μ, σ, a, b, c and S_p. Define $\psi = [1 + \sigma^2\mu^2]/2$.

Steps:

1. If either $S_p \leq \hat{S}_p \equiv \dfrac{\lambda_p\psi}{\mu[\mu - \lambda_p]}$ or $\dfrac{a}{c} \leq \dfrac{\lambda_p}{\mu^2}\psi$, then there does not exist a feasible solution. Assign $\lambda_s^* = 0$ and Stop. Else, go to the Step 2.

2. If $\dfrac{\mu - \lambda_p}{\mu\lambda_p} \leq \dfrac{a\lambda_p - c\psi}{2\mu\lambda_p^2 + c\psi(\mu + \lambda_p)}$, then assign $J_\ell = \infty$ and go to the Step 3. Else, find $\lambda_s^{(3)}$ the unique root of the cubic $\tilde{G}(\lambda_s)$ which lies in the interval $(0, \mu - \lambda_p)$ where

$$\tilde{G}(\lambda_s) \equiv 2\mu\lambda_s^3 - \left[a\mu + c\psi + 4\mu^2\right]\lambda_s^2 + 2\mu\left[a\mu + c\psi + \mu^2\right]\lambda_s - \mu\left[a\mu^2 - c\psi\lambda_p\right].$$

Calculate $J_\ell = \frac{\psi\lambda_3}{[\mu-\lambda_s^{(3)}][\mu-\lambda_3]}$ and define an interval $J = (J_\ell, \infty)$ where $\lambda_3 = \lambda_p + \lambda_s^{(3)}$. If $S_p \in J$, then assign $\lambda_s^* = \lambda_s^{(3)}$, $\beta^* = \infty$ and directly go to Step 6. Else, go to the Step 3.

3. If $\frac{a}{c} \leq \frac{\lambda_p(2\mu - \lambda_p)}{\mu(\mu - \lambda_p)^2}\psi$, then define an interval $J^- = (\hat{S}_p, J_\ell]$ when J_ℓ is finite otherwise take $J^- = (\hat{S}_p, \infty)$. Assign $\lambda_s^* = \frac{1}{2S_p}\left[S_p[2\mu - \lambda_p] + \psi - \sqrt{[S_p\lambda_p + \psi]^2 + 4\mu\psi S_p}\right]$, $\beta^* = \infty$ for $S_p \in J^-$ and directly go to Step 6. Else, go to the Step 4.

4. Find $\lambda_s^{(1)}$, the unique root of the cubic $G(\lambda_s)$ in the interval $(0,\ \mu - \lambda_p)$ with $\phi_0 = \mu - \lambda_p$ and

$$G(\lambda_s) = 2\mu\lambda_s^3 - [c\psi + \mu(a + 4\phi_0)]\lambda_s^2 + 2\phi_0[c\psi + \mu(a + \phi_0)]\lambda_s - a\mu\phi_0^2 + c\psi\lambda_p(\mu + \phi_0).$$

Calculate $I_\ell = \frac{\psi\lambda_1}{\mu[\mu-\lambda_p]}$ and $I_u = \frac{\psi\lambda_1}{[\mu-\lambda_s^{(1)}][\mu-\lambda_1]}$ where $\lambda_1 = \lambda_p + \lambda_s^{(1)}$.

5. Define intervals: $I^- = (\hat{S}_p, I_\ell)$, $I = [I_\ell, I_u)$ and $I^+ = [I_u, J_\ell]$ when J_ℓ is finite, otherwise take I^+ as $I^+ = [I_u, \infty)$.

   (a) If $S_p \in I^-$, then assign $\lambda_s^* = \frac{\mu[\mu-\lambda_p]S_p}{\psi} - \lambda_p$ and $\beta^* = 0$

   (b) If $S_p \in I$, then assign $\lambda_s^* = \lambda_s^{(1)}$ and

$$\beta^* = \begin{cases} \dfrac{[\mu - \lambda_1][\mu S_p[\mu - \lambda_p] - \psi\lambda_1]}{\psi\lambda_1^2 - \mu S_p\lambda_p[\mu - \lambda_1]} & \text{for } \dfrac{\psi\lambda_1}{\mu[\mu - \lambda_p]} \leq S_p \leq \dfrac{\psi\lambda_1}{\mu[\mu - \lambda_1]} \\[2ex] \dfrac{S_p\lambda_s^{(1)}[\mu - \lambda_1]}{\psi\lambda_1 - S_p[\mu - \lambda_s^{(1)}][\mu - \lambda_1]} & \text{for } \dfrac{\psi\lambda_1}{\mu[\mu - \lambda_1]} < S_p < \dfrac{\psi\lambda_1}{[\mu - \lambda_s^{(1)}][\mu - \lambda_1]} \end{cases}$$

   (c) If $S_p \in I^+$, then assign $\lambda_s^* = \frac{1}{2S_p}\left[S_p[2\mu - \lambda_p] + \psi - \sqrt{[S_p\lambda_p + \psi]^2 + 4\mu\psi S_p}\right]$ and $\beta^* = \infty$.

6. If given problem is feasible, the optimum assured service level to the secondary class customers is $S_s^* = W_s(\lambda_s^*, \beta^*)$ and the optimal unit admission price charged to the secondary class customers is $\theta^* = [a - cS_s^* - \lambda_s^*]/b$.

Figures 6-9 illustrate the possible intervals of S_p and objective function values at optimum within those intervals depending on the values of the input parameters.

4.2 Examples

We present numerical examples to illustrate the different possible cases of the algorithm depending upon the relative values of the input parameters.

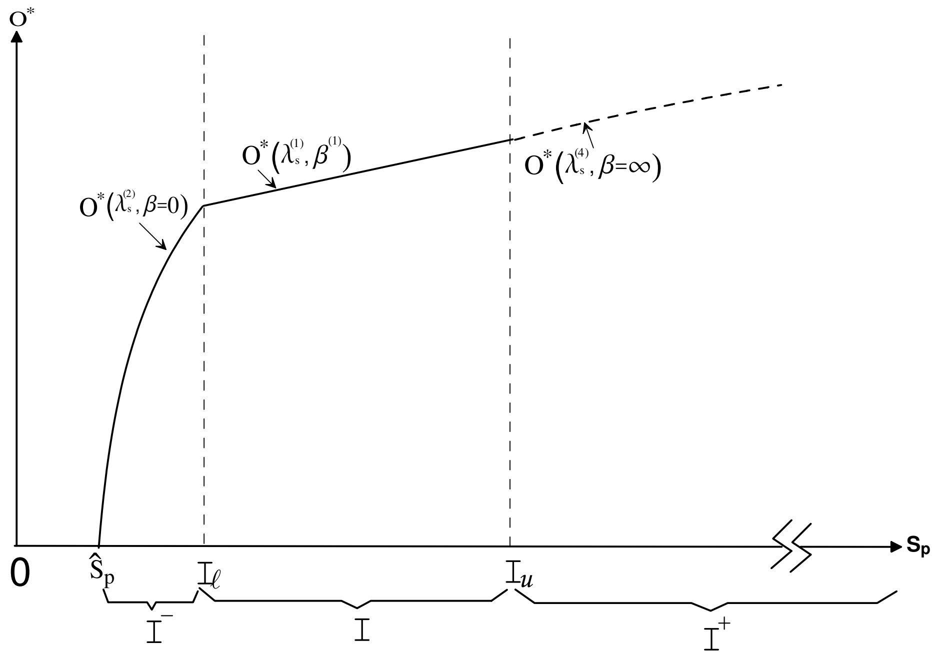


Figure 6: Optimal values of P0 and possible intervals of S_p whenever $\frac{a}{c} > \frac{\lambda_p(2\mu-\lambda_p)}{\mu(\mu-\lambda_p)^2}\psi$ and $\frac{\mu-\lambda_p}{\mu\lambda_p} \leq \frac{a\lambda_p - c\psi}{2\mu\lambda_p^2 + c\psi(\mu+\lambda_p)}$

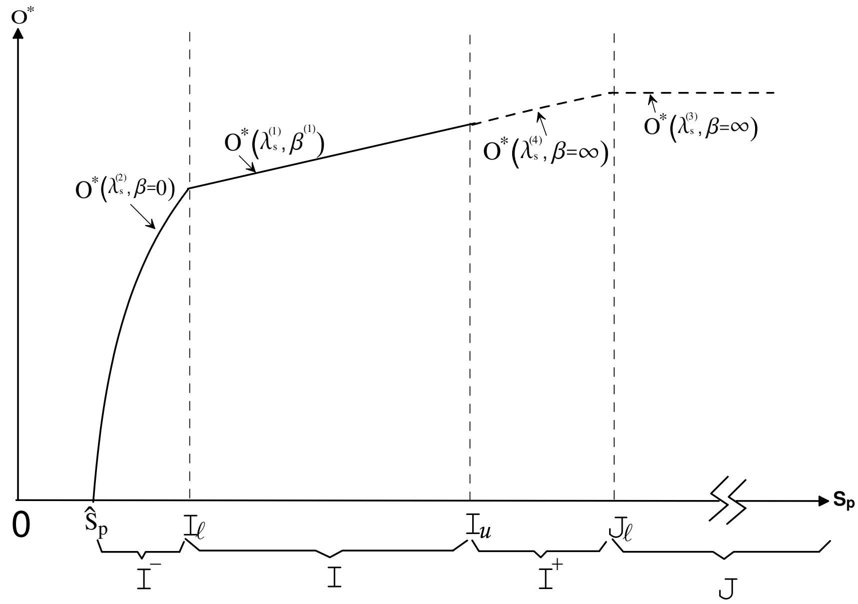


Figure 7: Optimal values of P0 and possible intervals of S_p whenever $\frac{a}{c} > \frac{\lambda_p(2\mu-\lambda_p)}{\mu(\mu-\lambda_p)^2}\psi$ and $\frac{\mu-\lambda_p}{\mu\lambda_p} > \frac{a\lambda_p - c\psi}{2\mu\lambda_p^2 + c\psi(\mu+\lambda_p)}$

Example 1: $\left(\frac{a}{c} > \frac{\lambda_p(2\mu-\lambda_p)}{\mu(\mu-\lambda_p)^2}\psi \text{ and } \frac{\mu-\lambda_p}{\mu\lambda_p} \leq \frac{a\lambda_p - c\psi}{2\mu\lambda_p^2 + c\psi(\mu+\lambda_p)}\right)$. Let us assume that the secondary class customers' demand function $\Lambda_s(\theta, S_s) = 100 - 0.2\theta - 0.1S_s$. Also, $\lambda_p = 8$ customers/hr, $\mu = 10$ customers/hr and $\sigma = 0.1$ hr/customer. We get $\psi = 1$, $\hat{S}_p = 0.4$ hr/customer, $\frac{a}{c} = 1000$ hr/customer, $\frac{\lambda_p}{\mu^2}\psi = 0.08$ hr/customer, $\frac{\lambda_p(2\mu-\lambda_p)}{\mu(\mu-\lambda_p)^2}\psi = 2.4$ hr/customer, $\frac{\mu-\lambda_p}{\mu\lambda_p} = 0.025$ hr/customer and $\frac{a\lambda_p - c\psi}{2\mu\lambda_p^2 + c\psi(\mu+\lambda_p)} = 0.624$ hr/customer. We note

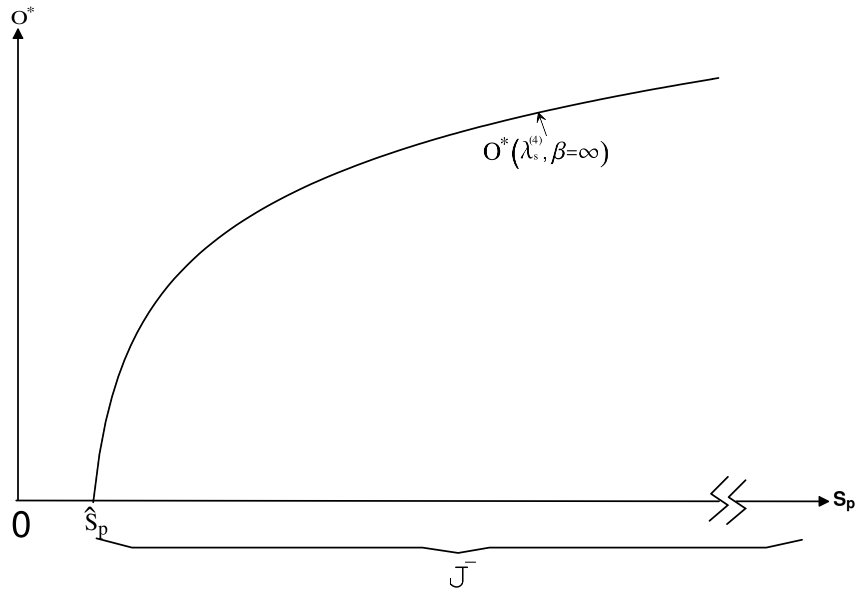


Figure 8: Optimal values of P0 and possible interval of S_p whenever $\frac{\lambda_p}{\mu^2}\psi < \frac{a}{c} \leq \frac{\lambda_p(2\mu-\lambda_p)}{\mu(\mu-\lambda_p)^2}\psi$ and $\frac{\mu-\lambda_p}{\mu\lambda_p} \leq \frac{a\lambda_p-c\psi}{2\mu\lambda_p^2+c\psi(\mu+\lambda_p)}$

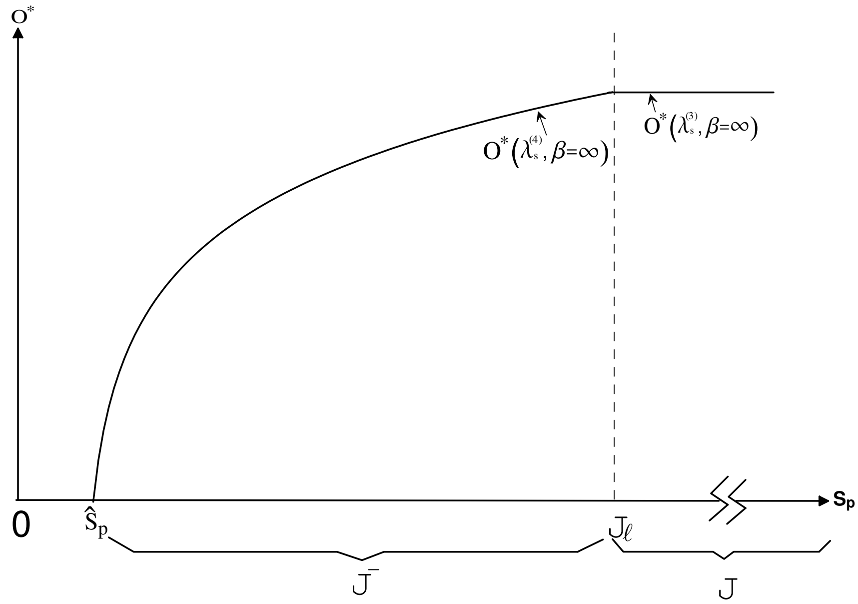


Figure 9: Optimal values of P0 and possible intervals of S_p whenever $\frac{\lambda_p}{\mu^2}\psi < \frac{a}{c} \leq \frac{\lambda_p(2\mu-\lambda_p)}{\mu(\mu-\lambda_p)^2}\psi$ and $\frac{\mu-\lambda_p}{\mu\lambda_p} > \frac{a\lambda_p-c\psi}{2\mu\lambda_p^2+c\psi(\mu+\lambda_p)}$

that a feasible solution will exist if $S_p > 0.40$ as $\frac{a}{c} > \frac{\lambda_p}{\mu^2}\psi$. As $\frac{\mu-\lambda_p}{\mu\lambda_p} < \frac{a\lambda_p-c\psi}{2\mu\lambda_p^2+c\psi(\mu+\lambda_p)}$ and $\frac{a}{c} > \frac{\lambda_p(2\mu-\lambda_p)}{\mu(\mu-\lambda_p)^2}\psi$, we directly go to Step 4. The calculations at Step 4 result in $\lambda_s^{(1)} = 1.898$, $I_\ell = 0.4949$ and $I_u = 11.97$. The intervals I^-, I and I^+ are $(0.4, 0.4949)$, $[0.4949, 11.97)$ and $[11.97, \infty)$ respectively. A few results corresponding to distinct values of S_p are presented in Table 1.

Example 2: $\left(\frac{a}{c} > \frac{\lambda_p(2\mu-\lambda_p)}{\mu(\mu-\lambda_p)^2}\psi \text{ and } \frac{\mu-\lambda_p}{\mu\lambda_p} > \frac{a\lambda_p-c\psi}{2\mu\lambda_p^2+c\psi(\mu+\lambda_p)}\right)$. Take demand function $\Lambda_s(\theta, S_s)$

Table 1: Representative results of Example 1

S_p	Priority β^*	Arrival rate λ_s^*	Price θ^*	Assured SL S_s^*	Revenue O^*
0.41	0	0.2	497.86	2.28	99.57
0.42	0	0.4	496.69	2.62	198.68
0.45	0	1	492.75	4.5	492.75
0.4949 $(=I_\ell)$	0	1.898	466.25	48.52	884.94
1	0.01	1.898	467.32	46.39	886.96
3	0.07	1.898	471.53	37.95	894.97
6	0.23	1.898	477.85	25.32	906.96
9.703	1	1.898	485.66	9.70	921.78
10	1.18	1.898	486.27	8.48	922.96
11	2.64	1.898	488.39	4.24	926.96
11.97 $(=I_u)$	∞	1.898	490.46	0.1222	930.87
13	∞	1.905	490.41	0.1223	934.65
14	∞	1.912	490.37	0.1226	937.82
15	∞	1.918	490.35	0.1227	940.58

as $4 - 0.2\theta - 0.1S_s$ in Example 1. This change results in $\frac{a}{c} = 30$ hr/customer and $\frac{a\lambda_p - c\psi}{2\mu\lambda_p^2 + c\psi(\mu+\lambda_p)} = 0.0186$ hr/customer. We note that $\frac{a}{c} > \frac{\lambda_p(2\mu-\lambda_p)}{\mu(\mu-\lambda_p)^2}\psi = 2.4$ and $\frac{\mu-\lambda_p}{\mu\lambda_p} = 0.025 > \frac{a\lambda_p - c\psi}{2\mu\lambda_p^2 + c\psi(\mu+\lambda_p)}$. A feasible solution will exist if $S_p > 0.40$ as $\frac{a}{c} > \frac{\lambda_p}{\mu^2}\psi = 0.08$. The calculations in Steps 2 and 4 of algorithm result in $\lambda_s^{(3)} = 1.493$, $\lambda_s^{(1)} = 1.002$, $I_\ell = 0.4501$, $I_u = 1.002$ and $J_\ell = 2.201$. The intervals I^-, I, I^+, J are $(0.4, 0.4501)$, $[0.4501, 1.002)$, $[1.002, 2.201]$ and $(2.201, \infty)$ respectively. A few results corresponding to distinct values of S_p are presented in Table 2.

Table 2: Representative results of Example 2

S_p	Priority β^*	Arrival rate λ_s^*	Price θ^*	Assured SL S_s^*	Revenue O^*
0.41	0	0.2	12.86	2.28	2.58
0.435	0	0.7	9.83	3.35	6.88
0.4501 $(= I_l)$	0	1.002	7.73	4.51	7.75
0.6	0.09	1.002	8.33	3.32	8.35
0.75	0.28	1.002	8.93	2.21	8.95
0.902	1	1.002	9.53	0.906	9.56
0.95	1.994	1.002	9.72	0.523	9.75
1.002 $(= I_u)$	∞	1.002	9.942	0.100	9.958
1.3	∞	1.196	8.965	0.105	10.727
1.6	∞	1.328	8.307	0.108	11.030
1.9	∞	1.422	7.835	0.110	11.141
2.201 $(= J_\ell)$	∞	1.493	7.481	0.112	11.166
2.5	∞	1.493	7.481	0.112	11.166
3	∞	1.493	7.481	0.112	11.166

Example 3: $\left(\frac{\lambda_p}{\mu^2}\psi < \frac{a}{c} \le \frac{\lambda_p(2\mu-\lambda_p)}{\mu(\mu-\lambda_p)^2}\psi \text{ and } \frac{\mu-\lambda_p}{\mu\lambda_p} \le \frac{a\lambda_p - c\psi}{2\mu\lambda_p^2 + c\psi(\mu+\lambda_p)}\right)$. Take demand function

$\Lambda_s(\theta, S_s)$ as $100-0.2\theta-400S_s$ in Example 1. This change results in $\frac{a}{c} = 0.25$ hr/customer, and $\frac{a\lambda_p - c\psi}{2\mu\lambda_p^2 + c\psi(\mu+\lambda_p)} = 0.0472$ hr/customer. Note that $\frac{\lambda_p}{\mu^2}\psi = 0.08 < \frac{a}{c} \leq \frac{\lambda_p(2\mu-\lambda_p)}{\mu(\mu-\lambda_p)^2}\psi = 2.4$ and $\frac{\mu-\lambda_p}{\mu\lambda_p} = 0.025 \leq \frac{a\lambda_p - c\psi}{2\mu\lambda_p^2 + c\psi(\mu+\lambda_p)}$. A feasible solution will exist if $S_p > 0.40$ as $\frac{a}{c} > \frac{\lambda_p}{\mu^2}\psi$. The interval J^- is $(0.4, \infty)$. A few results corresponding to distinct values of S_p are presented in Table 3.

Table 3: Representative results of Example 3

S_p	Priority β^*	Arrival rate λ_s^*	Price θ^*	Assured SL S_s^*	Revenue O^*
0.41	∞	0.034	338.61	0.081	11.47
1	∞	1.000	295.00	0.100	295.00
4	∞	1.707	257.34	0.117	439.38
8	∞	1.849	249.09	0.120	460.56
13	∞	1.906	245.70	0.1219	468.28
15	∞	1.918	244.96	0.1227	469.88
20	∞	1.938	243.74	0.1232	472.47

Example 4: $\left(\frac{\lambda_p}{\mu^2}\psi < \frac{a}{c} \leq \frac{\lambda_p(2\mu-\lambda_p)}{\mu(\mu-\lambda_p)^2}\psi \text{ and } \frac{\mu-\lambda_p}{\mu\lambda_p} > \frac{a\lambda_p - c\psi}{2\mu\lambda_p^2 + c\psi(\mu+\lambda_p)}\right)$. Take demand function $\Lambda_s(\theta, S_s)$ as $100-0.2\theta-550S_s$ in Example 1. This change results in $\frac{a}{c} = 0.182$ hr/customer and $\frac{a\lambda_p - c\psi}{2\mu\lambda_p^2 + c\psi(\mu+\lambda_p)} = 0.0224$ hr/customer. Note that $\frac{\lambda_p}{\mu^2}\psi = 0.08 < \frac{a}{c} \leq \frac{\lambda_p(2\mu-\lambda_p)}{\mu(\mu-\lambda_p)^2}\psi = 2.4$ and $\frac{\mu-\lambda_p}{\mu\lambda_p} = 0.025 > \frac{a\lambda_p - c\psi}{2\mu\lambda_p^2 + c\psi(\mu+\lambda_p)}$. Feasible solution will exist if $S_p > 0.40$ as $\frac{a}{c} > \frac{\lambda_p}{\mu^2}\psi$. The calculations in Steps 2 of algorithm result in $\lambda_s^{(3)} = 1.908$, and $J_\ell = 13.309$. The intervals J^- and J are $(0.4, 13.309]$ and $(13.309, \infty)$ respectively. A few results corresponding to distinct values of S_p are presented in Table 4.

Table 4: Representative results of Example 4

S_p	Priority β^*	Arrival rate λ_s^*	Price θ^*	Assured SL S_s^*	Revenue O^*
0.41	∞	0.034	278.14	0.081	9.42
1	∞	1.000	220.00	0.100	220.00
4	∞	1.707	169.55	0.117	289.48
8	∞	1.849	158.47	0.120	293.01
11	∞	1.889	155.26	0.121	293.31
13.309 ($= J_\ell$)	∞	1.908	153.76	0.122	293.34
14	∞	1.908	153.76	0.122	293.34
15	∞	1.908	153.76	0.122	293.34
20	∞	1.908	153.76	0.122	293.34

5 Nature of the optimal pricing and operating points

The optimum values of the decision variables of the constrained optimization problem, naturally depend on the prevailing QoS level to the primary class customers as well as on two of the three coefficients of secondary class customers' demand function. Also, the algorithm, defined in Section 4.1, divides the feasible region of S_p into intervals. Observe that either λ_s^* or β^* remain constant in those intervals. For example, the optimum queue discipline management parameter β^* remains unchanged in the intervals I^-, I^+ and J^- whereas the optimum mean arrival rate of the secondary class customers λ_s^* remains unchanged in the interval I. Section 5.1 elaborates on the qualitative nature of the optimal decisions. We also analyze the effect of prevailing QoS level of primary class customers S_p and coefficients of secondary class customers' demand function on the optimum decisions in Sections 5.2 and 5.3 respectively.

5.1 Key features of optimal operating points

The optimum decisions depend on the prevailing QoS level to the primary class customers S_p as well as on the ratio $\frac{a}{c}$ of the coefficients of secondary class customers' demand function. The optimal constrained resource sharing model is infeasible if this ratio $\frac{a}{c}$ is less than or equal to the threshold, $\frac{\lambda_p}{\mu^2}\psi$. If this ratio lies in a finite interval $(\frac{\lambda_p}{\mu^2}\psi, \frac{\lambda_p(2\mu-\lambda_p)}{\mu(\mu-\lambda_p)^2}\psi]$ which is to the right of above threshold, then, the optimum policy *always* assigns static high priority to the secondary class customers and allows the maximum possible arrival rate of secondary class customers that does not violate the prevailing QoS level to the primary class customers.

On the other hand, if the ratio $\frac{a}{c}$ to the right of above interval, then, the optimum policy depends on the prevailing QoS level to the primary class customers, S_p. The three intervals, I^-, I and I^+ correspond to low, moderate and high values of S_p. We note that lower the value of S_p, higher the QoS level to the primary class customers. When QoS level to primary class customers is high, (i.e., low S_p), then, the optimum operating policy assigns static high priority to the primary class customers and allows a maximum possible arrival rate of the secondary class customers as long as it does not violate the prevailing QoS level to the primary class customers. Similarly, when QoS level to primary class customers is low, (i.e., high S_p), then, the optimum operating policy assigns static high priority to the secondary class customers and allows a maximum possible arrival rate of the secondary class customers that does not violate the prevailing QoS level to the primary class customers. On the contrary, when QoS level to primary class customers is moderate, (i.e., moderate S_p), then, the optimum operating policy chooses a constant arrival rate of the secondary class customers for any value of S_p in that interval *but* employs a dynamic priority queue management scheme.

Also, we note below that across these intervals, both the pricing parameters, θ_s^* and S_s^* are non-linear but well behaved functions as are the pair of operating parameters β^* and λ_s^*.

5.2 Sensitivity analysis of optimal pricing w.r.t. S_p

We observed earlier that the optimum queue discipline management parameter $\beta^* = 0$ for any value of $S_p \in I^-$ and the optimum arrival rate of the secondary class customers λ_s^* is linearly increasing for $S_p \in I^-$ (from Corollary 2). Below, we study the effect of S_p on the contract parameters, i.e, optimum price θ^* and assured service level S_s^* in interval I^-.

Lemma 1. *In the interval I^-, the optimum price θ^* is a decreasing concave function and the optimum assured service level S_s^* is an increasing convex function of S_p.*

Proof. The optimum assured service level $S_s^* = W_s(\lambda_s^*, 0)$ for $S_p \in I^-$. We note that

$$\frac{\partial S_s^*}{\partial S_p} = \frac{\partial W_s}{\partial S_p} = \frac{\partial W_s}{\partial \lambda_s^*}\frac{\partial \lambda_s^*}{\partial S_p} \geq 0 \quad \text{and} \quad \frac{\partial^2 S_s^*}{\partial S_p^2} = \frac{\partial^2 W_s}{\partial {\lambda_s^*}^2}\left(\frac{\partial \lambda_s^*}{\partial S_p}\right)^2 + \frac{\partial W_s}{\partial \lambda_s^*}\frac{\partial^2 \lambda_s^*}{\partial S_p^2} \geq 0.$$

The above inequalities follow because $W_s(\lambda_s, \beta)$ is increasing convex function of λ_s and λ_s^* is linearly increasing function of S_p in the interval I^-. The optimal price charged to the secondary class customers $\theta^* = [a - cS_s^* - \lambda_s^*]/b$. We have

$$\frac{\partial \theta^*}{\partial S_p} = -\frac{1}{b}\left[c\frac{\partial S_s^*}{\partial S_p} + \frac{\partial \lambda_s^*}{\partial S_p}\right] \leq 0 \quad \text{and} \quad \frac{\partial^2 \theta^*}{\partial S_p^2} = -\frac{1}{b}\left[c\frac{\partial^2 S_s^*}{\partial S_p^2} + \frac{\partial^2 \lambda_s^*}{\partial S_p^2}\right] \leq 0.$$

Thus, θ^* is decreasing concave function of $S_p \in I^-$. □

Corollary 1 established that the optimum mean arrival rate of the secondary class customers λ_s^* remains unchanged in the interval I. Below, we study the effect of S_p on the optimal values of other decision variables in the interval I.

Lemma 2. *The optimum relative priority β^* and the optimum price θ^* are increasing convex functions of S_p in the interval I while the optimum assured service level S_s^* a decreasing concave function of S_p in the interval I.*

Proof. We observe that β^*, from Step 5-(b) of algorithm, is a continuous function of $S_p \in I$. The first order partial derivative of β^* with respect to S_p is

$$\frac{\partial \beta^*}{\partial S_p} = \begin{cases} \dfrac{\mu^2 \lambda_s^{(1)} \lambda_1 \psi(\mu - \lambda_1)}{[\psi \lambda_1^2 - \mu S_p \lambda_p(\mu - \lambda_1)]^2} & \text{for } \dfrac{\psi \lambda_1}{\mu[\mu - \lambda_p]} \leq S_p \leq \dfrac{\psi \lambda_1}{\mu[\mu - \lambda_1]} \\[2em] \dfrac{\lambda_1 \lambda_s^{(1)} \psi(\mu - \lambda_1)}{[\psi \lambda_1 - S_p(\mu - \lambda_s^{(1)})(\mu - \lambda_1)]^2} & \text{for } \dfrac{\psi \lambda_1}{\mu[\mu - \lambda_1]} < S_p < \dfrac{\psi \lambda_1}{[\mu - \lambda_s^{(1)}][\mu - \lambda_1]}. \end{cases}$$

We note that $\frac{\partial \beta^*}{\partial S_p} \geq 0$ as $\lambda_p > 0$, $\lambda_s^{(1)} > 0$ and $\lambda_p + \lambda_s^{(1)} < \mu$. Also, $\frac{\partial \beta^*}{\partial S_p}$ is continuous function of $S_p \in I$. The second order partial derivative of β^* with respect to S_p is

$$\frac{\partial^2 \beta^*}{\partial S_p^2} = \begin{cases} \dfrac{2\mu^3\lambda_p\lambda_s^{(1)}\lambda_1\psi(\mu-\lambda_1)^2}{[\psi\lambda_1^2 - \mu S_p\lambda_p(\mu-\lambda_1)]^3} & \text{for } \dfrac{\psi\lambda_1}{\mu[\mu-\lambda_p]} \le S_p \le \dfrac{\psi\lambda_1}{\mu[\mu-\lambda_1]} \\[2ex] \dfrac{2\lambda_1\lambda_s^{(1)}\psi(\mu-\lambda_s^{(1)})(\mu-\lambda_1)^2}{[\psi\lambda_1 - S_p(\mu-\lambda_s^{(1)})(\mu-\lambda_1)]^3} & \text{for } \dfrac{\psi\lambda_1}{\mu[\mu-\lambda_1]} < S_p < \dfrac{\psi\lambda_1}{[\mu-\lambda_s^{(1)}][\mu-\lambda_1]}. \end{cases}$$

We note that $\frac{\partial^2\beta^*}{\partial S_p^2} \ge 0$ as $\lambda_p > 0$, $\lambda_s^{(1)} > 0$ and $\lambda_p + \lambda_s^{(1)} < \mu$. Thus, β^* is an increasing convex function of $S_p \in I$. As $S_s^* = W_s(\lambda_s^*, \beta^*)$, we get,

$$\frac{\partial S_s^*}{\partial S_p} = \frac{\partial W_s}{\partial S_p} = \frac{\partial W_s}{\partial \beta^*}\frac{\partial \beta^*}{\partial S_p} \le 0 \quad \text{and} \quad \frac{\partial^2 S_s^*}{\partial S_p^2} = \frac{\partial^2 W_s}{\partial \beta^{*2}}\left(\frac{\partial \beta^*}{\partial S_p}\right)^2 + \frac{\partial W_s}{\partial \beta^*}\frac{\partial^2 \beta^*}{\partial S_p^2} \le 0.$$

The above inequalities follow because $W_s(\lambda_s, \beta)$ is a decreasing concave function of β and β^* is an increasing convex function of $S_p \in I$. The optimal price $\theta^* = [a - cS_s^* - \lambda_s^*]/b \equiv -f(S_s^*)$ if $S_p \in I$ as λ_s^* remains constant in this interval of S_p. Therefore, the optimum price θ^* is an increasing convex function of $S_p \in I$. □

Further, note that $\beta^* = \infty$ for S_p in the intervals I^+ and J^- while λ_s^* is increasing function of S_p in these intervals (Corollary 3). Below, we study effect of S_p on the optimum price θ^* and assured service level S_s^* in these intervals of S_p.

Lemma 3. *The optimum price θ^* is a decreasing function and the optimum assured service level S_s^* is an increasing function of S_p in both the intervals I^+ and J^-.*

Proof. As $S_s^* = W_s(\lambda_s^*, \infty)$, we have

$$\frac{\partial S_s^*}{\partial S_p} = \frac{\partial W_s}{\partial S_p} = \frac{\partial W_s}{\partial \lambda_s^*}\frac{\partial \lambda_s^*}{\partial S_p} \ge 0.$$

The above inequality follows because $W_s(\lambda_s, \beta)$ and λ_s^* are increasing functions of λ_s and S_p respectively. The optimal price $\theta^* = [a - cS_s^* - \lambda_s^*]/b$. We note that

$$\frac{\partial \theta^*}{\partial S_p} = -\frac{1}{b}\left[c\frac{\partial S_s^*}{\partial S_p} + \frac{\partial \lambda_s^*}{\partial S_p}\right] \le 0.$$

This implies that θ^* is a decreasing function of S_p. □

The results are summarized in the Table 5. The effect of S_p on the optimum revenue follows from the analysis done in Section 3.3. The optimal values remain constant in the interval J. We observe from Table 5 that the optimum contract parameters θ^* and S_s^* are different non-linear functions in the different intervals of S_p.

Table 5: Effect on optimal decisions with increase in S_p within an interval

Interval	Relative priority β^*	Arrival rate λ_s^*	Price θ^*	Assured service level S_s^*	Revenue O^*
I^-	Constant (0)	Increases (linear)	Decreases (concave)	Increases (convex)	Increases (concave)
I	Increases (convex)	Constant	Increases (convex)	Decreases (concave)	Increases (linear)
I^+ and J^-	Constant (∞)	Increases	Decreases	Increases	Increases (concave)

5.3 Sensitivity analysis of optimal decision variables w.r.t. demand function coefficients

We investigate the role of the two co-efficients of demand function, a and c, on the optimal choices of λ_s and β by looking at the dependence of the roots of cubics $G(\lambda_s)$ and $\tilde{G}(\lambda_s)$ on these co-efficients.

We established in Claim 5 that the increase in a results in increase of $\lambda_s^{(1)}$. This and similar arguments give the following results:

- The root of the cubic $G(\lambda_s)$, $\lambda_s^{(1)}$, is an increasing function of a and is a decreasing function of c.
- The root of the cubic $\tilde{G}(\lambda_s)$, $\lambda_s^{(3)}$, is increasing function of a and is a decreasing function of c.

Note that, $\frac{\partial I_\ell}{\partial \lambda_s^{(1)}} = \frac{\psi}{\mu[\mu-\lambda_p]} > 0$ and $\frac{\partial I_u}{\partial \lambda_s^{(1)}} = \frac{[\mu(\mu+\lambda_p)-\lambda_1^2]\psi}{[\mu-\lambda_s^{(1)}]^2[\mu-\lambda_1]^2} > 0$. The above inequalities follow as $\lambda_p \geq 0$, $\lambda_s^{(1)} > 0$ and $\lambda_1 = \lambda_p + \lambda_s^{(1)} < \mu$. Thus, the increase in $\lambda_s^{(1)}$ shifts I_ℓ and I_u to the right. Similarly, increase in $\lambda_s^{(3)}$ shifts J_ℓ to right as $\frac{\partial J_\ell}{\partial \lambda_s^{(3)}} > 0$. Further, we note that $\frac{\partial^2 I_\ell}{\partial {\lambda_s^{(1)}}^2} = 0$ and $\frac{\partial^2 I_u}{\partial {\lambda_s^{(1)}}^2} > 0$, i.e., $\frac{\partial I_\ell}{\partial \lambda_s^{(1)}}$ is constant while $\frac{\partial I_u}{\partial \lambda_s^{(1)}}$ is an increasing function of $\lambda_s^{(1)}$. The minimum of $\frac{\partial I_u}{\partial \lambda_s^{(1)}} = \frac{[\mu(\mu+\lambda_p)-\lambda_p^2]\psi}{\mu^2[\mu-\lambda_p]^2}$ at $\lambda_s^{(1)} = 0$. As,

$$\frac{\partial I_u}{\partial a} - \frac{\partial I_\ell}{\partial a} = \left[\frac{\partial I_u}{\partial \lambda_s^{(1)}} - \frac{\partial I_\ell}{\partial \lambda_s^{(1)}}\right]\frac{\partial \lambda_s^{(1)}}{\partial a} > \left[\frac{[\mu(\mu+\lambda_p)-\lambda_p^2]\psi}{\mu^2[\mu-\lambda_p]^2} - \frac{\psi}{\mu[\mu-\lambda_p]}\right]\frac{\partial \lambda_s^{(1)}}{\partial a}$$
$$= \left[\frac{(2\mu-\lambda_p)\lambda_p\psi}{\mu^2[\mu-\lambda_p]^2}\right]\frac{\partial \lambda_s^{(1)}}{\partial a} > 0,$$

the increase in demand coefficient a will widen the interval I. Note that the optimal policy uses delay dependent priority queue discipline in the interval I. Thus, the increase in a makes the delay dependent priority queue discipline to be used as part of optimal policy for a wider range of S_p values. Similarly, the increase in c narrows interval I and therefore, the delay dependent priority queue discipline is used as a part optimal policy for a smaller range of S_p values. Note that optimum admission rate remains constant over such intervals. Continuing with Example 1 described in Section 4.2, Tables 6(a) and 6(b) exhibit the effect on key parameters with variation in demand coefficients a and c respectively. We omit the infeasible values of $\lambda_s^{(1)}$ and $\lambda_s^{(3)}$ from the Tables 6(a) and 6(b).

Table 6: Effect of demand function co-efficient on key parameters

(a) Varying a; $c = 0.1$

a	$\lambda_s^{(1)}$	$\lambda_s^{(3)}$	I_ℓ	I_u
100	1.898	-	0.495	11.977
60	1.867	-	0.493	9.122
20	1.754	-	0.488	4.808
10	1.616	-	0.481	2.987
5	1.343	-	0.467	1.643
4	1.208	1.991	0.460	1.322
3	1.002	1.493	0.450	1.002
2	0.706	0.994	0.435	0.724
1	0.326	0.495	0.416	0.514

(b) Varying c; $a = 100$

c	$\lambda_s^{(1)}$	$\lambda_s^{(3)}$	I_ℓ	I_u
0.1	1.898	-	0.495	11.977
1	1.678	-	0.484	3.612
5	1.285	-	0.464	1.490
10	0.995	-	0.450	0.994
20	0.591	-	0.430	0.648
40	0.038	-	0.402	0.411
550	-	1.908	-	-
600	-	1.696	-	-
750	-	1.158	-	-

6 Discussion

We present closed form expressions for unit admission price for the secondary class customers and dynamic delay dependent queue discipline management parameter at a resource which is shared by two different classes of customers, primary class customers (existing customers) and secondary class customers (of new firms), each class being guaranteed that their mean queue lengths do not exceed certain values. We incorporate dynamic non-preemptive priority queue management discipline as a mechanism of admission control in this setting, to model the fact that the customers' demand function have to be sensitive to the admission price as well as to the assured service level. We find that in some cases, it is beneficial to offer a high priority to the secondary class customers while in other cases one has to employ the dynamic queueing discipline. We identify the regions of the input parameter space corresponding to these cases. We also present an extensive sensitivity analysis of these optimal decisions to various input parameters. We observe that these optimal decisions are non-linear functions of input parameters, in most cases. For example, the optimal unit admission price, θ^*, is concave decreasing in interval I^- of S_p and convex increasing in interval I of S_p but it is a decreasing function in interval I^+ of S_p.

We defined quality of service level as the expected waiting time in queue. It is possible to have waiting time distributions with large variances such that the expected values satisfy the specified service level. This may result in poor service level to a few customers due to large variances. We also know that even though the multi class queues remain stable under heavy traffic intensity, the resulting long queues are difficult to manage because of limited waiting space. Therefore, further studies can incorporate more demanding service levels such as variability in queue lengths or bounds on instances of unusually large delays via bounds on tail probabilities of waiting times. We assumed that the demand of secondary class customers is a linear function of the unit admission price and the assured service level. A similar analysis can be attempted with a non-linear demand function, e.g., log-linear Cobb-Douglas demand function (So and Song 1998). A subsequent study can also incorporate more than two classes of customers.

We assume certain values for the demand function coefficients and do not address the issues of estimating those coefficients. The optimal policies do depend on the demand function coefficients and therefore the accuracies of those values will be crucial. It will be interesting to develop a procedure to estimate the demand function coefficients, as the secondary class customers are firms which are new to the business or are currently using an alternate means for similar service. Also, it may be that the nature of the demand function, i.e, linear or non-linear, as well as the parameters defining those demand functions are known privately to the customers and remain unknown to the resource owner. Under such circumstances, the pricing scheme should also be incentive compatible, i.e., make customers reveal those private values to the resource owner, perhaps by suitable incentive payments (Mas-Colell *et al.* 1995). We note that such incentive compatible pricing schemes also have to satisfy the service level constraints.

Appendix

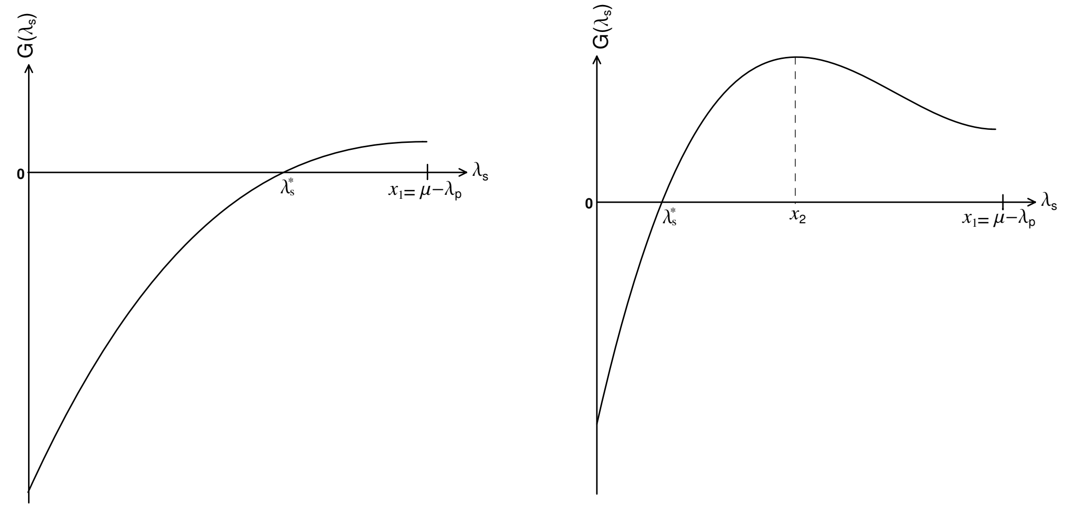
Proof of Claim 1. First, we observe that if $\frac{a}{c} > \frac{\lambda_p(2\mu-\lambda_p)}{\mu(\mu-\lambda_p)^2}\psi$ then the cubic $G(\lambda_s)$ has at least one positive root $\lambda_s \in (0, \mu - \lambda_p)$ as

$$G(0) = -a\mu\phi_0^2 + c\psi\lambda_p(\mu + \phi_0) < 0$$
$$G(\mu - \lambda_p) = c\psi\mu^2 > 0.$$

The first inequality follows from the assumption $\frac{a}{c} > \frac{\lambda_p(2\mu-\lambda_p)}{\mu(\mu-\lambda_p)^2}\psi$.

Next, we show that $G(\lambda_s)$ has only one root in the interval $(0, \mu - \lambda_p)$. The first derivative of the cubic is $G'(\lambda_s) = 2(\mu - \lambda_p - \lambda_s)(a\mu + c\psi + \mu^2 - \mu(\lambda_p + 3\lambda_s))$. So, the stationary points of the cubic $G(\lambda_s)$ are $x_1 = \mu - \lambda_p$ and $x_2 = \frac{1}{3\mu}\left[a\mu + c\psi + \mu(\mu - \lambda_p)\right]$. We note that $x_2 \gtreqless \mu - \lambda_p$ as $a\mu + c\psi \gtreqless 2\mu(\mu - \lambda_p)$. Let us first assume that $a\mu + c\psi \geq 2\mu(\mu - \lambda_p)$. Then, $G(\lambda_s)$ is an increasing function in the interval $[0, \mu - \lambda_p]$ as $G'(\lambda_s) \geq 6\mu\left(\mu - \lambda_p - \lambda_s\right)^2 > 0$. This scenario is illustrated by Figure 10(a). Therefore, the cubic $G(\lambda_s)$ has only one root in the interval $(0, \mu - \lambda_p)$ as it intersects the line $\lambda_s = 0$

only once in the interval $(0, \mu - \lambda_p)$.



(a) Case 1: $a\mu + c\psi \geq 2\mu(\mu - \lambda_p)$ (b) Case 2: $a\mu + c\psi < 2\mu(\mu - \lambda_p)$

Figure 10: Illustrative nature of cubic $G(\lambda_s)$ over the interval $[0, \mu - \lambda_p]$

Next, let us assume that $a\mu + c\psi < 2\mu(\mu - \lambda_p)$. The second derivative $G''(\lambda_s)$ evaluated at x_1 and x_2 results in $-2\left[a\mu + c\psi - 2\mu(\mu - \lambda_p)\right] > 0$ and $2\left[a\mu + c\psi - 2\mu(\mu - \lambda_p)\right] < 0$ respectively. This implies that x_1 and x_2 are points of relative minimum and maximum respectively. We also note that $G(\lambda_s)$ is an increasing function in the interval $[0, x_2]$ and a decreasing function in the interval $[x_2, x_1]$ and $G(x_2) > 0$. This scenario is illustrated by Figure 10(b). Again, the curve $G(\lambda_s)$ intersects the line $\lambda_s = 0$ only once in the interval $(0, \mu - \lambda_p)$, specifically in the interval $(0, x_2]$. Therefore, the cubic $G(\lambda_s)$ has only one root in the interval $(0, \mu - \lambda_p)$. □

Proof of Claim 2. Using equation (13), the equality $W_p(\lambda_s, \beta) = S_p$ results in either

$$\beta = \frac{[\mu - \lambda][\mu S_p[\mu - \lambda_p] - \psi\lambda]}{\psi\lambda^2 - \mu S_p \lambda_p [\mu - \lambda]} \tag{46}$$

or,

$$\beta = \frac{S_p \lambda_s [\mu - \lambda]}{\psi\lambda - S_p[\mu - \lambda_s][\mu - \lambda]}. \tag{47}$$

We observe that the obtained finite values of β (say $\bar{\beta}$) from equations (46) and (47) are feasible iff they satisfy $0 \leq \bar{\beta} \leq 1$ and $\bar{\beta} > 1$ respectively.

First we will analyze $\bar{\beta}$ obtained from equation (46). This value of $\bar{\beta}$ will be a positive number iff both numerator and denominator takes either positive or negative values simultaneously, i.e., either

$$\mu S_p[\mu - \lambda_p] - \psi\lambda \geq 0 \text{ and } \psi\lambda^2 - \mu S_p \lambda_p[\mu - \lambda] > 0$$

or

$$\mu S_p[\mu - \lambda_p] - \psi\lambda \leq 0 \text{ and } \psi\lambda^2 - \mu S_p \lambda_p[\mu - \lambda] < 0.$$

On simplification, we get, either

$$S_p \geq \frac{\psi\lambda}{\mu[\mu - \lambda_p]} \text{ and } S_p < \frac{\psi\lambda^2}{\mu\lambda_p[\mu - \lambda]} \tag{48}$$

or

$$S_p \leq \frac{\psi\lambda}{\mu[\mu - \lambda_p]} \text{ and } S_p > \frac{\psi\lambda^2}{\mu\lambda_p[\mu - \lambda]}. \tag{49}$$

We note that,

$$\frac{\psi\lambda^2}{\mu\lambda_p[\mu - \lambda]} - \frac{\psi\lambda}{\mu[\mu - \lambda_p]} = \frac{\psi\lambda\lambda_s}{\lambda_p[\mu - \lambda][\mu - \lambda_p]} > 0$$
$$\Rightarrow \frac{\psi\lambda^2}{\mu\lambda_p[\mu - \lambda]} > \frac{\psi\lambda}{\mu[\mu - \lambda_p]}.$$

From above inequality, we observe that it is impossible to have a positive S_p satisfying condition (49). Thus, the value of $\bar{\beta}$ obtained from equation (46) will be greater than or equal to zero iff following holds true:

$$\frac{\psi\lambda}{\mu[\mu - \lambda_p]} \leq S_p < \frac{\psi\lambda^2}{\mu\lambda_p[\mu - \lambda]}. \tag{50}$$

We need $\bar{\beta} \leq 1$ and hence from Equation (46) we have

$$\frac{[\mu - \lambda][\mu S_p[\mu - \lambda_p] - \psi\lambda]}{\psi\lambda^2 - \mu S_p \lambda_p[\mu - \lambda]} \leq 1.$$

On simplification, we get

$$S_p \leq \frac{\psi\lambda}{\mu[\mu - \lambda]}. \tag{51}$$

We note that,

$$\frac{\psi\lambda^2}{\mu\lambda_p[\mu - \lambda]} - \frac{\psi\lambda}{\mu[\mu - \lambda]} = \frac{\psi\lambda\lambda_s}{\mu\lambda_p[\mu - \lambda]} > 0$$
$$\Rightarrow \frac{\psi\lambda^2}{\mu\lambda_p[\mu - \lambda]} > \frac{\psi\lambda}{\mu[\mu - \lambda]}. \tag{52}$$

From inequalities (50) and (52), we infer that the obtained value of $\bar{\beta}$ from equation (46) will satisfy $0 \leq \bar{\beta} \leq 1$ iff

$$\frac{\psi\lambda}{\mu[\mu-\lambda_p]} \leq S_p \leq \frac{\psi\lambda}{\mu[\mu-\lambda]}. \tag{53}$$

Next, we will analyze $\bar{\beta}$ obtained from equation (47). The obtained value of $\bar{\beta}$ will be a finite positive number iff

$$\psi\lambda - S_p[\mu-\lambda_s][\mu-\lambda] > 0,$$

$$\text{i.e.,} \qquad S_p < \frac{\psi\lambda}{[\mu-\lambda_s][\mu-\lambda]}. \tag{54}$$

Also, we need $\bar{\beta} > 1$ and hence from Equation (47) we have

$$\frac{S_p\lambda_s[\mu-\lambda]}{\psi\lambda - S_p[\mu-\lambda_s][\mu-\lambda]} > 1.$$

On simplification, we get

$$S_p > \frac{\psi\lambda}{\mu[\mu-\lambda]} \tag{55}$$

From inequalities (54) and (55), we infer that the obtained value of $\bar{\beta}$ from equation (47) will satisfy $\bar{\beta} > 1$ iff

$$\frac{\psi\lambda}{\mu[\mu-\lambda]} < S_p < \frac{\psi\lambda}{[\mu-\lambda_s][\mu-\lambda]}. \tag{56}$$

□

Proof of the Claim 3. First, we observe that if $\frac{a}{c} > \frac{\lambda_p}{\mu^2}\psi$ then the cubic $\tilde{G}(\lambda_s)$ has at least one positive root $\lambda_s \in (0,\mu)$ as

$$\tilde{G}(0) = -\mu(a\mu^2 - c\psi\lambda_p) < 0$$
$$\tilde{G}(\mu) = c\psi\mu(\mu+\lambda_p) > 0.$$

The first inequality follows from the assumption $\frac{a}{c} > \frac{\lambda_p}{\mu^2}\psi$.

Next, we show that $\tilde{G}(\lambda_s)$ has only one root in the interval $(0,\mu)$. The first derivative of the cubic is $\tilde{G}'(\lambda_s) = 2(\mu-\lambda_s)(a\mu + c\psi + \mu^2 - 3\mu\lambda_s)$. So, the stationary points of the cubic $\tilde{G}(\lambda_s)$ are $\tilde{x}_1 = \mu$ and $\tilde{x}_2 = \frac{1}{3\mu}\left[a\mu + c\psi + \mu^2\right]$. We note that $\tilde{x}_2 \gtreqqless \mu$ as $a\mu + c\psi \gtreqqless 2\mu^2$. Let us first assume that $a\mu + c\psi \geq 2\mu^2$. Then, $\tilde{G}(\lambda_s)$ is an increasing function in the interval $[0,\mu]$ as $\tilde{G}'(\lambda_s) \geq 6\mu\left(\mu-\lambda_s\right)^2 > 0$. Therefore, the cubic $\tilde{G}(\lambda_s)$ has only one root in the interval $(0,\mu)$ as it intersects the line $\lambda_s = 0$ only once in the interval $(0,\mu)$.

Next, let us assume that $a\mu + c\psi < 2\mu^2$. The second derivative $\tilde{G}''(\lambda_s)$ evaluated at $\tilde{x}_1$ and $\tilde{x}_2$ are $-2\left[a\mu + c\psi - 2\mu^2\right] > 0$ and $2\left[a\mu + c\psi - 2\mu^2\right] < 0$ respectively. This implies that $\tilde{x}_1$ and $\tilde{x}_2$ are points of relative minimum and maximum respectively. We also note that $\tilde{G}(\lambda_s)$ is an increasing function in the interval $[0, \tilde{x}_2]$, a decreasing function in the interval $[\tilde{x}_2, \tilde{x}_1]$ and $\tilde{G}(\tilde{x}_2) > 0$. Again, the curve $\tilde{G}(\lambda_s)$ intersects the line $\lambda_s = 0$ only once in the interval $(0, \mu)$, specifically in the interval $(0, \tilde{x}_2]$. Therefore, the cubic $\tilde{G}(\lambda_s)$ has only one root in the interval $(0, \mu)$. □

Proof of Claim 4. The equality $\tilde{W}_p(\lambda_s) = S_p$ can be rewritten as a quadratic equation of λ_s.

$$Q(\lambda_s) \equiv S_p \lambda_s^2 - [S_p (2\mu - \lambda_p) + \psi] \lambda_s + \mu S_p (\mu - \lambda_p) - \psi \lambda_p = 0 \tag{57}$$

We observe that the quadratic $Q(\lambda_s)$ has at least one positive root $\lambda_s \in (0, \mu - \lambda_p)$ as

$$Q(0) = \mu S_p(\mu - \lambda_p) - \psi \lambda_p > 0$$
$$Q(\mu - \lambda_p) = -\psi \mu < 0.$$

The first inequality follows from the assumption that $S_p > \frac{\lambda_p \psi}{\mu[\mu - \lambda_p]}$. The roots, ω_1 and ω_2, of the quadratic equation (57) are

$$\omega_1 = \frac{1}{2S_p} \left[S_p [2\mu - \lambda_p] + \psi - \sqrt{[S_p \lambda_p + \psi]^2 + 4\mu\psi S_p} \right]$$
$$\omega_2 = \frac{1}{2S_p} \left[S_p [2\mu - \lambda_p] + \psi + \sqrt{[S_p \lambda_p + \psi]^2 + 4\mu\psi S_p} \right]$$

We note that both roots are real numbers. Also, $\omega_1 < \mu - \lambda_p$ and $\omega_2 > \mu + \frac{\psi}{S_p}$ as $[S_p \lambda_p + \psi]^2 + 4\mu\psi S_p > [S_p \lambda_p + \psi]^2$. Thus, ω_1 is the root of the quadratic equation (57) lying in the interval $(0, \mu - \lambda_p)$. □

Proof of Claim 5. To show the dependence of root of cubic $G(\lambda_s)$ with respect to a, we write the cubic as $G(\lambda_s, a)$. For $\lambda_s \in (0, \mu - \lambda_p)$, we observe that $\frac{\partial G(\lambda_s, a)}{\partial a} = -\mu(\mu - \lambda_p - \lambda_s)^2 < 0$, $\frac{\partial^2 G(\lambda_s, a)}{\partial a^2} = 0$. Thus, $G(\lambda_s, a)$ is a linearly decreasing function of a. Let a_1 and a_2 be the two different values of a such that $a_1 < a_2$. We note that $G(\lambda_s, a_1) > G(\lambda_s, a_2)$ for $\lambda_s \in (0, \mu - \lambda_p)$. Further, assume that $\frac{a_1}{c} > \frac{\lambda_p(2\mu - \lambda_p)}{\mu(\mu - \lambda_p)^2} \psi$. Under this assumption, we note from Claim 1 that the cubic $G(\lambda_s, a_1)$ and $G(\lambda_s, a_2)$ will have unique roots in the interval $(0, \mu - \lambda_p)$. Suppose $\lambda_s^{a_1}$ and $\lambda_s^{a_2}$ are the roots of the cubic $G(\lambda_s, a_1)$ and $G(\lambda_s, a_2)$ respectively. We note that $G(\lambda_s^{a_1}, a_1) = 0$ as $\lambda_s^{a_1}$ is the root of cubic $G(\lambda_s, a_1)$ and $G(\lambda_s^{a_1}, a_2) < 0$ as $G(\lambda_s, a_1) > G(\lambda_s, a_2)$ for a given $\lambda_s \in (0, \mu - \lambda_p)$. From Claim 1, we know that $G(0, a_2) < 0$. Therefore, the root of the cubic $G(\lambda_s, a_2)$ will lie right to $\lambda_s^{a_1}$, i.e., $\lambda_s^{a_1} < \lambda_s^{a_2}$. This implies that the root of the cubic $G(\lambda_s)$, $\lambda_s^{(1)}$, is increasing function of a. □

Some properties of $W_p(\lambda_s, \beta)$ and $W_s(\lambda_s, \beta)$

We observe from equations (13) and (14) that $W_p(\lambda_s, \beta)$ and $W_s(\lambda_s, \beta)$ always take finite positive values for $\lambda_p \geq 0$, $\lambda_s \geq 0$ and $\lambda_p + \lambda_s < \mu$. Also, $W_p(\lambda_s, \beta)$ and $W_s(\lambda_s, \beta)$ are continuous functions of β. Let us define

$$\begin{aligned} &\triangle_0 := \mu - \lambda, && \triangle_1 := \mu - \lambda_p [1-\beta], && \triangle_2 := \mu - \lambda [1-\beta], \\ &\triangle_3 := \mu - \lambda_s [1-1/\beta], && \triangle_4 := \mu - \lambda [1-1/\beta], && \triangle_5 := \mu + \lambda_p [1-1/\beta], \end{aligned}$$

and $\tilde{\beta} = 1 - 1/\beta$. Given $\lambda_p \geq 0$, $\lambda_s \geq 0$ and $\lambda_p + \lambda_s < \mu$, we note that $\triangle_0 > 0$, $\triangle_1, \triangle_2 > 0$ for $\beta \leq 1$ and $\triangle_3, \triangle_4, \triangle_5 > 0$ for $\beta \geq 1$.

Primary class customers' expected waiting time in queue $\boldsymbol{W_p(\lambda_s, \beta)}$: The first and second order partial derivatives of $W_p(\lambda_s, \beta)$ with respect to λ_s are

$$\begin{aligned} \frac{\partial W_p}{\partial \lambda_s} &= \psi \frac{[1-\beta] \triangle_0^2 + \beta \mu^2}{\mu \triangle_0^2 \triangle_1} \mathbb{1}_{\{\beta \leq 1\}} + \psi \frac{\triangle_0 \triangle_5 + \lambda \triangle_3}{\triangle_0^2 \triangle_3^2} \mathbb{1}_{\{\beta > 1\}} \\ \frac{\partial^2 W_p}{\partial \lambda_s^2} &= \frac{2\beta\mu\psi}{\triangle_0^3 \triangle_1} \mathbb{1}_{\{\beta \leq 1\}} + 2\psi \left[\frac{\lambda \tilde{\beta}^2}{\triangle_0 \triangle_3^3} + \frac{\mu \tilde{\beta}}{\triangle_0^2 \triangle_3^2} + \frac{\mu}{\triangle_0^3 \triangle_3} \right] \mathbb{1}_{\{\beta > 1\}}. \end{aligned}$$

We note that $\frac{\partial W_p}{\partial \lambda_s}, \frac{\partial^2 W_p}{\partial \lambda_s^2} \geq 0$ as $\lambda_p \geq 0$, $\lambda_s \geq 0$ and $\lambda_p + \lambda_s < \mu$. Therefore, $W_p(\lambda_s, \beta)$ is an increasing convex function of λ_s in the interval $[0,\ \mu - \lambda_p)$. The first and second order partial derivatives of $W_p(\lambda_s, \beta)$ with respect to β are

$$\begin{aligned} \frac{\partial W_p}{\partial \beta} &= \frac{\lambda_s \lambda \psi}{\triangle_0 \triangle_1^2} \mathbb{1}_{\{\beta \leq 1\}} + \frac{\lambda_s \lambda \psi}{\beta^2 \triangle_0 \triangle_3^2} \mathbb{1}_{\{\beta > 1\}} \\ \frac{\partial^2 W_p}{\partial \beta^2} &= -\frac{2\lambda_p \lambda_s \lambda \psi}{\triangle_0 \triangle_1^3} \mathbb{1}_{\{\beta \leq 1\}} - \frac{2[\mu - \lambda_s] \lambda_s \lambda \psi}{\beta^3 \triangle_0 \triangle_3^3} \mathbb{1}_{\{\beta > 1\}}. \end{aligned}$$

We note that $\frac{\partial W_p}{\partial \beta} \geq 0$ and $\frac{\partial^2 W_p}{\partial \beta^2} \leq 0$ $\lambda_p \geq 0$, $\lambda_s \geq 0$ and $\lambda_p + \lambda_s < \mu$. Therefore, $W_p(\lambda_s, \beta)$ is an increasing concave function of $\beta \geq 0$.

The diagonal elements of Hessian matrix $H[W_p(\lambda_s, \beta)]$, $\frac{\partial^2 W_p}{\partial \lambda_s^2}$ and $\frac{\partial^2 W_p}{\partial \beta^2}$, have opposite signs, i.e., the Hessian matrix is indefinite. This implies that $W_p(\lambda_s, \beta)$ is neither a convex nor a concave function of (λ_s, β) where $\lambda_s \in [0, \mu - \lambda_p)$ and $\beta \geq 0$.

Secondary class customers' expected waiting time in queue $\boldsymbol{W_s(\lambda_s, \beta)}$: The first and second order partial derivatives of $W_s(\lambda_s, \beta)$ with respect to λ_s are

$$\frac{\partial W_s}{\partial \lambda_s} = \frac{\mu \psi}{\triangle_0^2 \triangle_1} \mathbb{1}_{\{\beta \leq 1\}} + \psi \left[\frac{\lambda}{\beta \triangle_0^2 \triangle_3} + \frac{\triangle_4 \triangle_5}{\mu \triangle_0 \triangle_3^2} \right] \mathbb{1}_{\{\beta > 1\}}$$

$$\frac{\partial^2 W_s}{\partial \lambda_s^2} = \frac{2\mu\psi}{\triangle_0^3\triangle_1}\mathbb{1}_{\{\beta\le 1\}} + 2\psi\left[\frac{\tilde{\beta}\triangle_4\triangle_5}{\mu\triangle_0\triangle_3^3} + \frac{\triangle_5}{\beta\triangle_2^2\triangle_3^2} + \frac{\lambda}{\beta\triangle_2^2\triangle_3}\right]\mathbb{1}_{\{\beta>1\}}$$

We note that $\frac{\partial W_s}{\partial \lambda_s}, \frac{\partial^2 W_s}{\partial \lambda_s^2} \ge 0$ as $\lambda_p \ge 0$, $\lambda_s \ge 0$ and $\lambda_p + \lambda_s < \mu$. Therefore, $W_s(\lambda_s, \beta)$ is an increasing convex function of λ_s in the interval $[0,\ \mu - \lambda_p)$. The first and second order partial derivatives of $W_s(\lambda_s, \beta)$ with respect to β are

$$\frac{\partial W_s}{\partial \beta} = -\frac{\lambda_p\lambda\psi}{\triangle_0\triangle_1^2}\mathbb{1}_{\{\beta\le 1\}} - \frac{\lambda_p\lambda\psi}{\beta^2\triangle_0\triangle_3^2}$$

$$\frac{\partial^2 W_s}{\partial \beta^2} = \frac{2\lambda_p^2\lambda\psi}{\triangle_0\triangle_1^3}\mathbb{1}_{\{\beta\le 1\}} + \frac{2\left[\mu - \lambda_s\right]\lambda_p\lambda\psi}{\beta^3\triangle_0\triangle_3^3}\mathbb{1}_{\{\beta>1\}}$$

We note that $\frac{\partial W_s}{\partial \beta} \le 0$ and $\frac{\partial^2 W_s}{\partial \beta^2} \ge 0$ as $\lambda_p \ge 0$, $\lambda_s \ge 0$ and $\lambda_p + \lambda_s < \mu$. Therefore, $W_s(\lambda_s, \beta)$ is a decreasing convex function of $\beta \ge 0$.

References

Bazaraa, M. S., H. D. Sherali and C. M. Shetty (1993), *Nonlinear programming: Theory and Algorithms*, New York: John Wiley.

Hall, J. M., P. K. Kopalle and D. F. Pyke (2002), Static and dynamic pricing of excess capacity in a make-to-order environment, Working paper no. 2004-01, Tuck School of Business, Dartmouth, NH. `http://ssrn.com/abstract=485702`.

Kanet, J. J. (1982), A mixed delay dependent queue discipline, *Operations Research*, **30**(1), pp. 93–96.

Kleinrock, L. (1964), A delay dependent queue discipline, *Naval Research Logistics Quarterly*, **11**, pp. 329–341.

Kleinrock, L. (1976), *Queueing Systems, Vol-II, Computer applications*, New York: John Wiley.

Mas-Colell, A., M. D. Whinston and J. R. Green (1995), *Microeconomic Theory*, New York: Oxford University Press.

Palaka, K., S. Erlebacher and D. Kropp (1998), Lead-time setting, capacity utilization, and pricing decisions under lead-time dependent demand, *IIE Transactions*, **30**, pp. 151–163.

Pekgun, P., P. M. Griffin and P. Keskinocak (2008), Coordination of marketing and production for price and leadtime decisions, *IIE Transactions*, **40**, pp. 12–30.

Ray, S. and E. Jewkes (2004), Customer lead time management when both demand and price are lead time sensitive, *European Journal of Operational Research*, **153**, pp. 769–781.

Sinha, S. K., N. Rangaraj and N. Hemachandra (2008), A model for service level based pricing of shared resources at container depots, in: *Proc. of the International Conference on Transportation System Studies (ICOTSS-08)*, University of Mumbai, Mumbai.

So, K. C. and J. Song (1998), Price, delivery time guarantees and capacity selection, *European Journal of Operational Research*, **111**, pp. 28–49.

Wolff, R. W. (1989), *Stochastic Modeling and the Theory of Queues*, Englewood Cliffs, New Jersey: Prentice Hall.