Optimal Service Capacities in a Competitive Multiple-Server Queueing Environment

Wai-Ki Ching[1], Sin-Man Choi[1], and Min Huang[2]

[1] Advanced Modeling and Applied Computing Laboratory,
Department of Mathematics,
The University of Hong Kong
{wching,kellyci}@hkusua.hku.hk
[2] College of Information Science and Engineering,
Northeastern University, Shenyang, 110004, China
Key Laboratory of Integrated Automation of Process Industry,
Ministry of Education, Shenyang, 110004, China
mhuang@mail.neu.edu.cn

Abstract. The study of economic behavior of service providers in a competition environment is an important and interesting research issue. A two-server queueing model has been proposed in Kalai et al. [11] for this purpose. Their model aims at studying the role and impact of service capacity in capturing larger market share so as to maximize the long-run expected profit. They formulate the problem as a two-person strategic game and analyze the equilibrium solutions. The main aim of this paper is to extend the results of the two-server queueing model in [11] to the case of multiple servers. We will only focus on the case when the queueing system is stable.

Keywords: Markovian Queueing Systems, n-server Queue, Nash Equilibrium, Competition.

1 Introduction

The problem of finding the optimal strategy and control policy of a queueing system is a traditional mathematical problem and has been well studied in the literature, see for instance [2,9,10,11,12,17]. In an optimal control problem, it usually involves making decisions on system parameters such as the system service capacity and number of servers in the system under a specified cost structure (convex or concave). Here service capacity is an important competitive factor in the design of a service system, for example, in the areas of telecommunication networks [6] data transmission systems [11] and Vendor-Managed Inventory (VMI) system [3,16]. In particular, the current development in supply chain management emphasizes the coordination and integration of inventory and transportation logistics [4,18]. VMI is a supply chain initiative where the distributor is responsible for all decisions regarding the selection of retailers or agents. This creates a competitive environment for the agents and retailers to compete in the market [14].

J. Zhou (Ed.): Complex 2009, Part I, LNICST 4, pp. 66–77, 2009.
© ICST Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2009

Kalai et al. [11] studied a strategic game of two servers competing for their market shares through determining their service capacities. A Markovian queueing system of two servers is used in their model and analysis. Markovian queueing systems are popular tools for modeling servicing systems as they are mathematically tractable [6,7] when compared to the non-Markovian queueing systems. The problem is then analyzed using game theory [15]. Game theory is a popular and promising approach [1,5] for the captured problem. They classified the Nash equilibria into three different cases concerning the cost function and the revenue per customer. The waiting time is finite in one of these cases and there is a unique symmetric equilibrium. Although their model is simple, it brings in two important concepts. The first one is the "competitive game of servers" and the second one is "the market share of a server in a multi-server facility". Furthermore, they also report that when the marginal cost of providing service is "high", there is a unique symmetric equilibrium and the total service capacity is less than the mean demand rate. In such a case, each server actually behaves as if it were a monopolist. Competition therefore has no effect and this leads to an undesirable situation. On the other hand, when the marginal cost of providing service is "low", a unique symmetric equilibrium exists and the total service capacity is greater than the mean demand rate. In this paper, we will extend the model in [11] by allowing the number of servers to be more than two. In particular, we are interested in the case when the total service capacity is greater than the mean demand rate.

The remainder of the paper is structured as follows. In Section 2, we will give a brief review on the two-server queueing system discussed in [11] and the analytic results therein. We then present our multiple-server queueing system and also our analysis on the system performance in Section 3. A numerical demonstration is given in Section 4 for the case of a 3-server queueing systems. Finally concluding remarks are given to address further research issues in Section 5.

2 A Review on the Two-Server Queueing System

The service system studied in Kalai et al [11] consists of two independently operated servers. Customers arrive according to a Poisson process of rate λ and the service times are assumed to follow the exponential distribution. Each of the server i operates independently and determines its own service capacity μ_i so as to maximize its own profits. The cost to operate at service capacity μ is $c(\mu)$. Here the operating cost function $c(.)$ is assumed to an increasing and strictly convex function, i.e., both $c'(\mu)$ and $c''(\mu)$ are both positive and an example of such a function is $c(\mu) = \mu^2$.

The servers earn a fixed amount R for each unit of service rendered. The queueing system consists of a single First-In-First-Out queue. If a customer arrives when both servers are idle, the customer will be assigned to either server with equal likelihood. No server is allowed to be idle when at least one customer in the system. If a customer arrives when one server is idle and the other is busy, he/she will be assigned to the idle server. In the following subsections, we will present briefly the main results obtained in [11] concerning the two-server queueing system.

2.1 The System Steady-State Probability Distribution

If Server i $(i = 1, 2)$ chooses service capacity μ_i and such that

$$\mu_1 + \mu_2 > \lambda \tag{1}$$

the system has a steady-state probability distribution. We remark that condition (1) is a necessary and sufficient condition for the Markovian queueing system to be stable or to have steady-state probability distribution. Let P_n be the probability that there are n customers in the system; P_{10} be the probability that server 1 is busy and server 2 is idle; P_{01} be the probability that server 2 is busy and server 1 is idle. By studying the balanced equations of the queueing system, we have the following results:

$$P_0 = \frac{1-\rho}{1-\rho+\frac{\lambda(\mu_1+\mu_2)}{2\mu_1\mu_2}} \quad \text{and} \quad P_{10} = \frac{\lambda P_0}{2\mu_1} \quad \text{and} \quad P_{01} = \frac{\lambda P_0}{2\mu_2} \tag{2}$$

where

$$\rho = \frac{\lambda}{(\mu_1 + \mu_2)} \tag{3}$$

is the system load. Moreover, we also have

$$P_1 = P_{10} + P_{01} \quad \text{and} \quad P_n = \rho^{n-1} P_1 \quad n = 2, 3, \ldots \tag{4}$$

2.2 The Market Share

Computing the market share of Server i is equivalent to computing the mean number of customers per time unit that enter service with Server i. Using the results in Section 2.1, if $\mu_1 + \mu_2 > \lambda$, the mean number of customers per time unit that enter service with Server 1 is

$$P_0 \frac{\lambda}{2} + P_{01}\lambda + P_3\mu_1 + P_4\mu_1 + \ldots \tag{5}$$

and that with Server 2 is

$$P_0 \frac{\lambda}{2} + P_{10}\lambda + P_3\mu_2 + P_4\mu_2 + \ldots . \tag{6}$$

We then divide by the mean number of customers per time unit that enter service, i.e., λ, to obtain the *market share* of Server i. Thus the fraction of all customers served by Server $i(i = 1, 2)$, is given by

$$\alpha_i(\mu_1, \mu_2) = \frac{\lambda\mu_i^2 + \mu_1\mu_2(\mu_1 + \mu_2)}{\lambda(\mu_1 + \mu_2)^2 + 2\mu_1\mu_2(\mu_1 + \mu_2 - \lambda)}. \tag{7}$$

2.3 The Profit Function

Given the market shares of the servers in Section 2.2, the profit function $\pi_i(\mu_1, \mu_2)$ for Server $i \in \{1, 2\}$, the expected profit per time unit earned by Server i, is then given by

$$\pi_i(\mu_1, \mu_2) = \begin{cases} R\lambda\alpha_i(\mu_1, \mu_2) - c(\mu_i) & \text{if } \mu_1 + \mu_2 > \lambda \\ R\mu_i - c(\mu_i) & \text{if } \mu_1 + \mu_2 \leq \lambda. \end{cases} \tag{8}$$

Here $c(\mu)$ is the cost of providing service at capacity μ and R is the revenue per customer served.

2.4 The Nash Equilibrium of the Queueing System

Kalai et al. [11] considered the situation as a two-person strategic game and they found that finite waiting times exist at equilibrium if and only if

$$c'(\frac{\lambda}{2}) < \frac{R}{2}. \tag{9}$$

Moreover, if this condition is satisfied, then a unique equilibrium exists in which both servers select the same service capacity $\mu_c = \mu_1 = \mu_2$ such that

$$c'(\mu_c) = \frac{R\lambda^2}{2\mu_c(2\mu_c + \lambda)}. \tag{10}$$

3 The General Multiple-Server Queueing System

In this section, we extend the two-server queueing system studied in [11] to a general n-server queueing system. The arrival process of customers is assumed to be a Poisson process. In this queueing system, arriving customers wait in a single First-In-First-Out (FIFO) queue if all servers are busy. No server is allowed to be idle when there is at least one customer in the queueing system. If a customer arrives when more than one server is idle, the customer is assigned to any of the idle servers with equal likelihood. Once a server completes the service of a customer, the first customer in the queue, if any, is assigned to the server. Each server i may choose its own service capacity μ_i, and its service time follows the exponential distribution with mean $1/\mu_i$. The servers earn a revenue of R per customer served, and each of them incurs a cost of $c(\mu)$ to operate at service capacity μ, where $c(.)$ is an increasing and strictly convex function, i.e., both $c'(.)$ and $c''(.)$ are both positive.

In the following subsections, we present some important properties of the multiple-server queueing system through the propositions. The proofs of the propositions are omitted but can be found in [8].

3.1 The Steady-State Distribution of the Queueing System

Given the service capacities $\mu_1, \ldots, \mu_n$ and the mean demand rate λ, suppose $\sum_{i=1}^{n} \mu_i > \lambda$. This condition is to guarantee that the queueing system is stable and the system steady-state probability distribution exists. We would like to obtain the steady-state probability distribution of the number of customers in the system. Let us give the following definitions. Let P_i be the steady-state probability of having i customers in the system, where $i = 0, 1, 2, \ldots$. Also let $P_{\mathbf{s}}$, where $\mathbf{s} = (s_1, s_2, \ldots, s_n)$ and $s_i = 0$ or 1, be the steady-state probability of having s_i customers at Server i. We note that by definition

$$P_k = \sum_{\{\mathbf{s} | s_1 + \ldots + s_n = k\}} P_{\mathbf{s}} \quad \text{for} \quad k = 0, 1, \ldots, n. \tag{11}$$

We establish the equations governing the steady-state probabilities. The equations can be obtained by equating the incoming rate and outgoing rate at each of the state. For $s_i = 0, 1$ and $\sum_{i=1}^{n} s_i \neq n$, we have

$$(\sum_{\{i | s_i = 1\}} \mu_i + \lambda) P_{(s_1, s_2, \ldots, s_n)} = \sum_{\{i | s_i = 0\}} \mu_i P_{(s_{-i}, s_i = 1)} + \sum_{\{i | s_i = 1\}} \frac{\lambda P_{(s_{-i}, s_i = 0)}}{|\{j | s_j = 0\}| + 1}. \tag{12}$$

where (s_{-i}, s_i') denotes $(s_1, \ldots, s_{i-1}, s_i', s_{i+1}, \ldots, s_n)$. When $s_i = 0$ for all i this gives

$$\lambda P_{(0,0,\ldots,0)} = \mu_1 P_{(1,0,\ldots,0)} + \mu_2 P_{(0,1,0,\ldots,0)} + \cdots + \mu_n P_{(0,\ldots,0,1)}. \tag{13}$$

For the states with at least n customers we have

$$(\sum_{i=1}^{n} \mu_i + \lambda) P_{(1,1,\ldots,1)} = \sum_{i=1}^{n} \mu_i P_{n+1} + \sum_{i=1}^{n} \lambda P_{(s_{-i} = \mathbf{1}, s_i = 0)} \tag{14}$$

and

$$(\sum_{i=1}^{n} \mu_i + \lambda) P_k = (\sum_{i=1}^{n} \mu_i) P_{k+1} + \lambda P_{k-1} \text{ for } k = n+1, n+2, \ldots. \tag{15}$$

We note that these two equations together are equivalent to

$$(\sum_{i=1}^{n} \mu_i + \lambda) P_k = (\sum_{i=1}^{n} \mu_i) P_{k+1} + \lambda P_{k-1} \text{ for } k = n, n+1, \ldots. \tag{16}$$

We also have the normalization equation

$$\sum_{i=0}^{\infty} P_i = 1. \tag{17}$$

It can be shown by direct verification that the solution is given by the following proposition.

Proposition 1. *We have*

$$P_{(s_1,s_2,\ldots,s_n)} = \frac{(n-k)!\lambda^k P_0}{n!\prod_{\{i|s_i=1\}}\mu_i} \quad \textit{where} \quad k = s_1 + s_2 + \ldots + s_n > 0 \tag{18}$$

and

$$P_k = \rho_{k-n} P_n \quad \textit{for} \quad k > n \tag{19}$$

and

$$P_0 = \left(1 + \sum_{k=1}^{n-1} \frac{(n-k)!\lambda^k (\sum_{i_1<i_2<\ldots<i_{n-k}} \mu_{i_1}\mu_{i_2}\ldots\mu_{i_{n-k}})}{n!\mu_1\mu_2\ldots\mu_n} + (\frac{1}{1-\rho})\frac{\lambda^n}{n!\mu_1\mu_2\ldots\mu_n}\right)^{-1}. \tag{20}$$

The steady-state probability distribution describes the long-run behavior of the system. Each of these probabilities P_k represents the long-run proportion of time that there are k customers in the system. They are essential in studying how each server determines its strategy to maximize its long-run profit. In the next subsection, we will write the market share of each server in terms of these probabilities and obtain an expression for the market share.

3.2 The Market Share of Each Server

We derive the market share of each server from the steady-state distribution. We note that when $\sum_{j=1}^{n}\mu_j \leq \lambda$, i.e., customers arrive at least as fast as the servers can serve them, the steady-state probability distribution does not exist and the queue is infinite. In this case, each server receives customers at its service capacity in the long run. Otherwise, $\sum_{j=1}^{n}\mu_j > \lambda$ and all customers will be served. Each server only receives a fraction of the arriving customers, at a rate lower than its service capacity. The server's profit thus depends on the fraction of all customers it serves, i.e. its market share.

When $k(1 \leq k \leq n)$ servers are idle, customers arrive at a rate of λ and an arriving customer is served by any one of the k idle servers with equal likelihood. Each of these idle servers therefore receives customers at a rate of λ/k. On the other hand, when all servers are busy with at least one customer waiting in the system, each of the busy servers i receives a new customer when it completes the service for a customer, i.e. at a rate of its service capacity μ_i.

To obtain the market share, we find the expected value of the server's rate of receiving customers in different states of the systems, taking expectation over the steady-state probabilities. In the following, we give the formula for the market share for an individual server.

Proposition 2. *If* $\sum_{j=1}^{n}\mu_j > \lambda$*, the market share of Server* i*,* $\alpha_i(\mu_1,\mu_2,\ldots,\mu_n)$ *is given by*

$$\frac{\mu_i\left[\sum_{k=0}^{n-1} k!\lambda^{n-k-1}\left(\sum_{j_1<j_2<\ldots<j_k, j_p\neq i\,\forall p} \mu_{j_1}\mu_{j_2}\ldots\mu_{j_k}\right) + \lambda^{n-1}\left(\frac{\rho}{1-\rho}\right)\right]}{\sum_{k=1}^{n} k!\lambda^{n-k}(\sum_{j_1<j_2<\ldots<j_k}\mu_{j_1}\mu_{j_2}\ldots\mu_{j_k}) + \frac{\lambda^n}{1-\rho}}. \tag{21}$$

As we focus on the case when the mean demand rate is less than the total service rate, the market share is directly tied to the profit of a server. Before formulating the profit function of a server, we state the following two propositions related to the partial derivatives of the market share α_i with respect to μ_i. These will be useful in determining the Nash equilibrium of the system when we considered the system as a n-player strategic game.

Proposition 3. *Suppose that* $\sum_{j=1}^{n} \mu_j > \lambda$ *then* $\frac{\partial \alpha_i(\mu_1,\mu_2,\ldots,\mu_n)}{\partial \mu_i} > 0$. *Furthermore, when* $\mu_i \to \infty$, *we have* $\frac{\partial \alpha_i(\mu_1,\ldots,\mu_n)}{\partial \mu_i} \to 0$.

Proposition 4. *Suppose that* $\sum_{j=1}^{n} \mu_j > \lambda$, *then* $\frac{\partial^2 \alpha_i(\mu_1,\mu_2,\ldots,\mu_n)}{\partial \mu_i^2} < 0$.

Propositions 3 and 4 together mean that the market share α_i is increasing and concave with respect to μ_i $(i = 1, 2, \ldots, n)$.

3.3 The Profit Function

Here we proceed to find out the profit function of an individual server, which represents the server's profit per time unit in the long run. There are two cases to be considered. Suppose that $\sum_{j=1}^{n} \mu_j > \lambda$, Server i receives customers at a rate of $\lambda\alpha_i(\mu_1, \mu_2, \ldots, \mu_n)$. When $\sum_{j=1}^{n} \mu_j \le \lambda$, Server i receives customer at a rate of μ_i. In both cases, Server i incurs a cost of $c(\mu_i)$. Therefore similar to [11], the profit function of Server i is given by

$$\pi_i(\mu_1, \mu_2, \ldots, \mu_n) = \begin{cases} R\lambda\alpha_i(\mu_1, \mu_2, \ldots, \mu_n) - c(\mu_i) & \text{if} \quad \displaystyle\sum_{j=1}^{n} \mu_j > \lambda \\ R\mu_i - c(\mu_i) & \text{if} \quad \displaystyle\sum_{j=1}^{n} \mu_j \le \lambda \end{cases} \tag{22}$$

Each of the servers aims to maximize its long-run profit when determining its service capacity. Therefore, how a server's profit changes with its service capacity (when other servers' capacities remain unchanged) is important in characterizing the server's decision. By proposition 3 and 4, we readily obtain the following proposition describing the properties of the profit function π_i with respect to μ_i.

Proposition 5. *For* $i = 1, 2, \ldots, n$, *for each fixed* $\lambda > 0$ *and* $\mu_j > 0$ *where* $j \neq i$, *the function* $\pi_i(\mu_1, \mu_2, \ldots, \mu_n)$ *is continuous and strictly concave in* μ_i.

The continuity and concavity of the profit function ensure that the first-order condition is a sufficient condition for a value of μ_i to maximize the profit function.

3.4 The Nash Equilibrium of the Queueing System

Since servers' decisions of their service capacities would affect the profit of each other, we model the situation as an n-player strategic game, in which each server i chooses its service capacity μ_i to maximize its profit π_i. Here we discuss the Nash equilibrium of the system. In the two-server model in [11], a unique symmetric equilibrium is found in the case when the total demand rate is less than the total service rate. In our analysis, we will show that, similar to the two-server case, when the marginal cost is low enough, there is a unique equilibrium, in which all servers choose the same service capacities. In the following, we will first look at how the profit of Server i changes with its service capacity when all other servers choose the same service capacities.

Proposition 6. *For $\mu_c > \lambda/n$,*

$$\frac{\partial}{\partial\mu_i}\alpha_i(\mu_1,\mu_2,\ldots,\mu_n)\bigg|_{\mu_1=\mu_2=\ldots=\mu_n=\mu_c} = \frac{\lambda}{n^2\mu_c^2}\left[1-\frac{\lambda^{n-1}}{\sum_{k=0}^{n-1}(k+1)!\binom{n-1}{k}\lambda^{n-k-1}\mu_c^k}\right]$$

which is decreasing in μ_c. Also, we have

$$\lim_{\mu_c\to(\lambda/n)^+}\frac{\partial}{\partial\mu_i}\alpha_i(\mu_1,\mu_2,\ldots,\mu_n)\bigg|_{\mu_1=\mu_2=\ldots=\mu_n=\mu_c} = \frac{n-1}{n\lambda}$$

and

$$\lim_{\mu_c\to\infty}\frac{\partial}{\partial\mu_i}\alpha_i(\mu_1,\mu_2,\ldots,\mu_n)\bigg|_{\mu_1=\mu_2=\ldots=\mu_n=\mu_c} = 0.$$

It should be noted that proposition 6 implies that for $\mu_c > \lambda/n$, we have

$$\frac{\partial}{\partial\mu_i}\alpha_i(\mu_1,\mu_2,\ldots,\mu_n)\bigg|_{\mu_1=\mu_2=\ldots=\mu_n=\mu_c} < \frac{n-1}{n\lambda}.$$

We also note that the partial derivative in proposition 6 gives the marginal benefit Server i gets by unilaterally deviating from a service capacity μ_c commonly chosen by all servers.

The following proposition gives the Nash equilibrium of the game, which represents the decision of the servers on their service capacities in the long run.

Proposition 7. *If $(n-1)R/n > c'(\lambda/n)$ then there is a unique equilibrium where $\mu_1 = \mu_2 = \ldots = \mu_n = \mu_c$ and μ_c is the unique solution that satisfies $\mu_c > \lambda/n$ and*

$$R\lambda\frac{\partial}{\partial\mu_i}\alpha_i(\mu_1,\mu_2,\ldots,\mu_n)\bigg|_{\mu_1=\mu_2=\ldots=\mu_n=\mu_c} = c'(\mu_c). \tag{23}$$

i.e.,

$$R\left(\frac{\lambda}{n\mu_c}\right)^2\left[1-\frac{\lambda^{n-1}}{\sum_{k=0}^{n-1}(k+1)!\binom{n-1}{k}\lambda^{n-k-1}\mu_c^k}\right] = c'(\mu_c). \tag{24}$$

If $(n-1)R/n \le c'(\lambda/n)$ then the system has no equilibrium in which the expected waiting time is finite.

We note that from the proposition, we have $\mu_c > \lambda/n$ and so the expected waiting times are finite. This means that we know that if the marginal cost of serving $1/n$ of all customers is less than $(n-1)/n$ of the revenue received per customer, there is a unique symmetric equilibrium with finite waiting times.

For equation (23) to hold, it means that the marginal benefit Server i gets by unilaterally deviating from a service capacity μ_c commonly chosen by all servers must be equal to the marginal cost to do so. In this case, Server i does not benefit from changing its service capacity. Mathematically, the first-order condition for π_i holds. From the concavity of π_i obtained in proposition 5, we know that choosing μ_c as the service capacity maximizes the profit for Server i.

Since the servers share the same cost function and the same profit function with respect to their own service capacities, the condition for which the marginal benefit equals the marginal cost is identical for all servers when they choose the same service capacities. The proposition asserts that there is only one value of μ_c which satisfies the condition, and that this symmetric equilibrium is the unique equilibrium of the system.

This proposition shows that, given the arrival rate of customer λ, the number of servers n and the revenue per customer R, all servers will choose the same service capacity given by equation (24) in the long run if the condition

$$\frac{(n-1)R}{n} > c'(\frac{\lambda}{n}) \tag{25}$$

is satisfied. The proposition is useful for determining the minimum value of revenue per customer R for which the system will have a finite-waiting time equilibrium.

When $n = 2$, Propositions 6 and 7 reduce to the results in [11]. It is worth noting that as n increases, $(n-1)R/n$ increases and $c'(\lambda/n)$ decreases. Therefore, the minimum value of R required for the existence of a finite waiting-time equilibrium decreases as n increases. An increase in the number of servers causes competition to become more intense. Thus the minimum revenue per customer needed to achieve an equilibrium with finite waiting times becomes lower.

4 A Numerical Example on Three-Server Queueing System

In this section, we present a numerical example for the case of a three-server queueing system, i.e., $n = 3$. Here we assume the cost function takes the following form:

$$c(\mu) = \mu^2 \tag{26}$$

and the condition for the queueing system to be stable

$$\mu_1 + \mu_2 + \mu_3 > \lambda. \tag{27}$$

We note that $c'(\mu) > 0$ and $c''(\mu) > 0$ for $\mu > 0$. Thus $c(\mu)$ is strictly increasing and strictly convex.

We first give the steady-state probability distribution of the system. The following result comes from Proposition 1 in Section 3.1. We have

$$P_0 = \frac{1-\rho}{(1-\rho)\left(1+\frac{\lambda(\mu_1\mu_2+\mu_1\mu_3+\mu_2\mu_3)}{2\mu_1\mu_2\mu_3}\right)+\frac{\lambda^2(\mu_1+\mu_2+\mu_3)}{6\mu_1\mu_2\mu_3}},$$

$$P_{(0,0,1)} = \frac{\lambda P_0}{3\mu_3}, \quad P_{(0,1,0)} = \frac{\lambda P_0}{3\mu_2}, \quad P_{(1,0,0)} = \frac{\lambda P_0}{3\mu_1},$$

$$P_{(0,1,1)} = \frac{\lambda^2 P_0}{6\mu_2\mu_3}, \quad P_{(1,0,1)} = \frac{\lambda^2 P_0}{6\mu_1\mu_3}, \quad P_{(1,1,0)} = \frac{\lambda^2 P_0}{6\mu_1\mu_2},$$

and

$$P_k = \rho^{k-2}P_2 \quad \text{for} \quad k > 2$$

where

$$P_2 = P_{(0,1,1)} + P_{(1,0,1)} + P_{(1,1,0)}.$$

Moreover, we have

$$\alpha_i(\mu_1,\mu_2,\mu_3) = \frac{\mu_i\left[\lambda^2+\lambda(\mu_j+\mu_l)+2\mu_j\mu_l+\frac{\lambda^3}{\mu_i+\mu_j+\mu_l-\lambda}\right]}{\lambda^2(\mu_i+\mu_j+\mu_l)+2\lambda(\mu_i\mu_j+\mu_i\mu_l+\mu_j\mu_l)+6\mu_i\mu_j\mu_l+\frac{\lambda^3(\mu_i+\mu_j+\mu_l)}{\mu_i+\mu_j+\mu_l-\lambda}}.$$

where $j, l \in \{1,2,3\}$ and i, j, l are distinct. Now we have

$$\left.\frac{\partial}{\partial\mu_i}\alpha_i(\mu_1,\mu_2,\mu_3)\right|_{\mu_1=\mu_2=\mu_3=\mu_c} = \frac{2\lambda(2\lambda+3\mu_c)}{9\mu_c(\lambda^2+4\mu_c\lambda+6\mu_c^2)}.$$

If $2R/3 > c'(\lambda/n) = 2\lambda/3$, i.e., $R > \lambda$ then there is a unique symmetric equilibrium where $\mu_1 = \mu_2 = \mu_3 = \mu_c$ and μ_c is the unique solution that satisfies $\mu_c > \lambda/3$ and

$$\left[\frac{2\lambda^2(2\lambda+3\mu_c)}{9\mu_c(\lambda^2+4\mu_c\lambda+6\mu_c^2)}\right]R = c'(\mu_c) = 2\mu_c$$

i.e.,

$$54\mu_c^4 + 36\lambda\mu_c^3 + 9\lambda^2\mu_c^2 - 3R\lambda^2\mu_c - 2R\lambda^3 = 0.$$

5 Concluding Remarks

In this paper, we extend the analytic results of the two-server queueing system discussed in [11] to an n-server queueing system. To extend our study to the incentive aspect of the queueing system is our future work.

In fact, a service system of two servers coordinated by one central agency was studied by Gilbert and Weng [12]. The principal-agent relationship [13] between the central agency and the servers was studied, from the principal's perspective. It is of interest whether the allocation policy with a *separate queue* or that with a *common queue* would allow the coordinator to control waiting times at a lower cost. The service system studied in [12] consists of two independently operated servers coordinated by one central agency. Again customers arrive according to a Poisson process and the service times are assumed to follow an exponential distribution. Each of the server operates independently and determines its own service capacity so as to maximize its individual profits. The coordinating agency determines a fixed amount R, the compensation to the servers for each unit of service rendered, to induce a desirable service capacity. The coordinating agency's goal is to minimize its cost to maintain expected sojourn time below a given level. It was found that the servers have a weaker incentives to increase their service capacities in common queue systems than in separate queue systems. In many cases, the competition incentive effects can more than offset the risk-pooling benefits of a common queue. In particular, cases with small permissible waiting times or not severe diseconomies on increasing capacity favor the separate queue system.

The queueing system discussed in this paper corresponds to the common queue with n servers. Therefore the results obtained here are ready to apply to generalize the models and conclusions addressed in [12].

Acknowledgment. Ching is supported in part by Hong Kong RGC Grant No. 7017/07P and the HKU Strategic Research Funding on Computational Sciences and CRCG Grants of the University of Hong Kong. Huang is supported in part by the National Natural Science Foundation of China (Project no. 70671020, 70721001, 70431003, 60673159), the Program for New Century Excellent Talents in University (Project no.NCET-05-0295,NCET-05-0289), Specialized Research Fund for the Doctoral Program of Higher Education (20070145017, ,20060145012). The authors would like to thank the anonymous referees for their helpful comments and corrections.

References

1. Altman, E.: Non-zero-sum Stochastic Games in Admission, Service and Routing Control in Queueing Systems. Queueing Systems Theory Appl. 23, 259–279 (1996)
2. Andradotir, S., Ayhan, H., Down, D.: Server Assignment Policies for Maximizing the Steady-State Throughput of Finite Queueing Systems. Manag. Sci. 47, 1421–1439 (2001)
3. Ben-Daya, M., Hariga, M.: Integrated Single Vendor Single Buyer Model with Stochastic Demand and Variable Lead Time. International Journal of Production Economics 92, 75–80 (2004)
4. Bernstein, F., Chen, F., Federgruen, A.: Coordinating Supply Chains with Simple Pricing Schemes: The Role of Vendor-Managed Inventories. Manag. Sci. 52, 1483–1492 (2006)
5. Ching, W.: On Convergence of Asynchronous Greedy Algorithm with Relaxation in Multiclass Queueing Environment. IEEE Communication Letters 3, 34–36 (1999)
6. Ching, W.: Iterative Methods for Queuing and Manufacturing Systems. Springer Monographs in Mathematics. Springer, London (2001)
7. Ching, W., Ng, M.: Markov Chains: Models, Algorithms and Applications. International Series on Operations Research and Management Science. Springer, New York (2006)

8. Ching, W., Choi, S., Huang, M.: Optimal Service Capacity in a Multiple-server Queueing System: A Game Theory Approach (preprint) (2008), `http://hkumath.hku.hk/papers/~wkc/cchpaper1.pdf`
9. Crabill, C., Gross, D., Magazine, M.: A Classified Bibliography of Research on Optimal Control of Queues. Oper. Res. 25, 219–232 (1977)
10. El-Taha, M., Maddah, B.: Allocation of Service Time in a Multiserver System. Manag. Sci. 52, 623–637 (2006)
11. Kalai, E., Kamien, M., Rubinovitch, M.: Optimal Service Speeds in a Competitive Environment. Manag. Sci. 38(8), 1154–1163 (1992)
12. Gilbert, S., Weng, Z.: Incentive Effects Favor Nonconsolidating Queues in a Service System: The Principal-Agent Perspective. Manag. Sci. 44(12), 1662–1669 (1998)
13. Laffont, J., Martimort, D.: The Theory of Incentives: the Principal-agent Model. Princeton University Press, Princeton (2002)
14. Mishra, B., Raghunathan, S.: Retailer vs. Vendor-Managed Inventory and Brand Competition. Manag. Sci. 50, 445–457 (2004)
15. Morries, P.: Introduction to Game Theory. Springer, New York (1994)
16. Tai, A., Ching, W.: A Quantity-time-based Dispatching Policy for a VMI System. In: Gervasi, O., Gavrilova, M.L., Kumar, V., Laganá, A., Lee, H.P., Mun, Y., Taniar, D., Tan, C.J.K. (eds.) ICCSA 2005. LNCS, vol. 3483, pp. 342–349. Springer, Heidelberg (2005)
17. Teghem, J.: Control of the Service Process in a Queueing System. Euro. J. of Oper. Res. 23, 141–158 (1986)
18. Thomas, D.: Coordinated Supply Chain Management. European Journal of Operational Research 94, 1–15 (1996)