Whether or Not to Allow Preoccupation of Seats in Cafeteria Queues
카페테리아 대기행렬 내 좌석 선점 허용 문제 연구

Seung Bum Soh(First Author)
School of Business, Yonsei University
(*sbsoh@yonsei.ac.kr*)

Our queueing model is motivated by cafeterias where food and seats are provided. Cafeterias vary according to the policy they impose on the preoccupation of seats: customers are either allowed to secure the seats before the food or they should first obtain the food and next take the seats. We analyze the problem of selecting the policy that results in higher social welfare using a stylized fluid model. First, the dynamics of the system are identified in both of the policies. These results are used to construct stationary states of the system. Finally, the social welfares in the stationary states from both of the policies are compared and it turns out that the social welfares are identical and the choice of the policy is irrelevant of the efficiency.

Key Words: Queueing system, Behavioral queueing theory, Fluid approximation, Cafeteria, Service system

Ⅰ. Introduction

In usual restaurants and cafeterias, two main kinds of service are provided - food and seats. Customers need seats to eat food. Typical processes customers go through in restaurants are as follows: be seated, order food, receive the food and eat. In cafeterias, the process may slightly be different: order food, receive the food, be seated and eat. However, it is not always the case and some customers first occupy seats before they go to obtain the food. In that case, the sequence a customer goes through become similar to the one in restaurants.

However, occupying seats before obtaining the food is sometimes considered as an impolite behavior especially in self-serviced cafeterias in universities (Saini, 2017). On the surface, it is a readily justifiable convention since the seats once occupied but not yet used

Submission Date: 08. 22. 2023 Accepted Date: 09. 11. 2023

Copyright 2023 THE KOREAN ACADEMIC SOCIETY OF BUSINESS ADMINISTRATION
This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0, which permits unrestricted, distribution, and reproduction in any medium, provided the original work is properly cited.

can be considered as a temporarily wasted resource. The objective of this paper is to analyze this convention and observe whether the behavior of occupying the seat first, if everybody follows it, is detrimental to the social welfare. This question has not answered in previous literature to the best of our knowledge.

We first provide a novel setting to study this problem. In the mathematical model, customers receive a positive utility (which we call 'reward') from receiving the service. The net utility from the service is this reward subtracted by the disutility from waiting by adopting a widely accepted fact that customers take care of the time they spend waiting (see, Cho & Kim, 2007; Park, 2007). The waiting comes from two sources: waiting for the seat and waiting for the food. The customers are assumed to calculate the net utility from the system and decide whether to join the system or not. The customers who do not join the queue receive the net utility normalized to 0.

In the first policy, customers are not allowed to occupy the seat before obtaining the food. Only customers who have received the service can be seated. In the second policy, customers are allowed to occupy the seat first. We assume that once it is allowed, all the incoming customers will take the seat first. We first describe the dynamics of the systems under two policies and obtain the performances of the two policies and compare them.

The typical queueing perspective is used in this analysis. However, obtaining the full exact description of the system is possibly a daunting task, which might be reflected in the lack of papers that deal with this question. Broadly, there are two ways to overcome the difficulty that lies in the exact analyslis of the system. One is to construct simulations (Park et al., 1999). The other way is to study approximated version and that is what we do in this paper. There are two broad kinds of approximations in queueing theory - fluid approximation and diffusion approximation. The fluid approximation can be thought of as a queueing version of the law of large numbers, while the diffusion approximation is a queueing version of the central limit theorem. In the fluid approximation, most of the stochasticities are intentionally ignored, while the first moments of bigger scale parameters are discussed. In the diffusion approximation, the second moments are also incorporated in the model.

We use the fluid approximation to describe the system and to find the stationary state that enables a comparison of the performances. We also use the framework that actively reflects the fact that customers are utility maximizer. The decision making process is explicitly presented in this framework, and some of the behavior of customers are not given as an exogenous one but an endogenous one. Especially, we assume that the decision of whether to join the system or to balk are result of utility maximization decision.

In our model, it turns out that the performances of the two policies - disallowing the occupation of the seat before getting the food and allowing it - are the same in terms of social welfare. The identity of the social welfare is an unanticipated result, especially to the people who perceive no-preoccupation policy to be superior in terms of efficiency. Hence, there is no ground to prohibit occupation at least in our framework. However, there may be other reasons that make it inefficient and we will also discuss these in the final section.

Ⅱ. Literature review

We are studying a queueing network policy selection problem that is motivated by those in cafeterias. There are a few works that dealt with cafeterias. June & Jain(2010) resembles our work in that it studied a cafeteria queue using a fluid limit. However, the property of cafeteria queue June & Jain(2010) focus on was that it has a starting time and customers can arrive before it. It is a property that a cafeteria queue shares with a concert queue. The customer behavior on what time to arrive at the queue before the system opens is derived. Weber & Weiss(1994) and Füßler et al.(2019) also studied cafeteria queues and focused on the property that a cafeteria serves different foods. In Weber & Weiss(1994), a cafeteria is depicted as a system that provides multiple dishes. Customers make one line and they may or may not demand the dishes. Phenomena that arise in that kind of queueing network are studied. In Füßler et al.(2019), a single waiter is serving different food and customers demand different set of dishes. Customer sequencing and waiter scheduling problem were studied using mixed integer programming. Lastly, Stout Jr.(1995) and Lee & Lambert (2006) used simulation to study cafeteria queues. Among the above, none has focused on the property of cafeteria that we handled.

Our work can be regarded as a paper that studies queuing systems using economic framework. Naor(1969) is deemed as the origin of the long stretch of works that are in this category. See Hassin & Haviv(2003) and Hassin(2013) for the surveys of this kind of works.

Especially, we are using fluid limits to facilitate the discussion. Fluid limits were used in various contexts in queueing literature and there are huge amount of works that used fluid limits on descriptive queueing model. Typically, the sequence of original queueing systems is constructed and the fluid model in the limit is conjectured. The main parts of those works determined whether the sequence of the queueing systems converges to the given fluid model.

In this literature review, we will review the works that applied fluid limit in the economic framework which include reward-cost structures or customer behavior. Rajagopal et al. (1995) is one of the first paper that uses fluid queue in the economic perspective. The typical reward-cost structure was incorporated in the fluid limits. Fluid models are used in more conventional queueing problem of revenue or social welfare optimization as in Maglaras & Zeevi(2005) where two class of services- guaranteed and best-effort – are provided and optimal prices are decided. Moreover, higher tractability of fluid models have allowed analysis of various unconventional queueing systems. Allon & Gurvich(2010) and studied the equilibrium behavior of competing service providers. Both fluid scale and diffusion scale were used for the analysis. Allon et al.(2017) used fluid models to analyze the role of moderating firm in skill marketplace. Ata et al.(2017) applied it in a complex kidney transplantation problem. Afeche et al.(2017) put together various objectives and activities of large call centers that were typically studied individually in one model by using fluid models. Recently, on-off server problem with strategic customers were studied mainly with the fluid models. See Economou & Manou(2016), Liu et al.(2021), Wang & Xu(2021) and Logothetis et al. (2022).

Ⅲ. Model description

We first provide notations that can also cover dynamic cases, while our main analysis is on the stationary state. The customers are assumed to arrive with the rate of $\lambda(t)$ where t denotes the time. Later in this paper, the incoming rate of the customers are given as a decision from utility maximizer. We focus on two kinds of service in cafeterias – obtaining the food and eating the food on the seats – and the time that it takes to get those services are incorporated in the model. All other miscellaneous constituents, for example, going back and forth to the tables and serving areas are not considered in the model. Food is obtained with the rate of μ_1 and the individual customer will finish eating with the rate of μ_2. Since only the first moment remains in fluid models, if the incoming rate is $\lambda(t)$, it not only means that the average number of the customers that enter the system is $\lambda(t)$ but also exactly the number of incoming customers in the unit time is $\lambda(t)$, i.e., typical randomness is not incorporated in the model. It also applies to the service rates μ_1 and μ_2.

There are two kinds of queues in our model – one before the food and one before the table. Depending on the policies, customers may first queue up before the food and second before the tables or first before the tables and next before the food. Regardless of the policies, we denote the queue for the food at time t as $Q_1(t)$ and the one for the table at time t as $Q_2(t)$. Note that the customers may first wait in $Q_2(t)$ before they wait in $Q_1(t)$ depending on the policy. In fluid models, the number of customers are allowed to be non-integer and also the queue lengths are defined as real numbers.

The number of tables is denoted as B. The tables may be occupied or not and the occupying customers may be dining or waiting for the food. $p_1(t)$ is the proportion of the tables at time t that are occupied by the customers who are still waiting for the food. Note that $p_1(t)$ can be positive only in the policy that allows occupation of the table before obtaining the food. $p_2(t)$ is the proportion of the tables at time t that are occupied by the customers who are dining. Note that $p_1(t)+p_2(t) \le 1$ holds for all t. Since we are constructing a fluid model and atomic property of the customers are ignored, we assume $p_1(t)$ and $p_2(t)$ can take any value if they are non-negative and $p_1(t)+p_2(t) \le 1$ holds.

We assume that service rate of the whole tables are $Bp_2(t)\mu_2$. Hence, if all the tables are occupied with the dining customers, i.e., $p_2(t)=1$ the service rate is calculated as $Bp_2(t)\mu_2 = B\mu_2$. The assumption is motivated by $M/M/N$ queue where the mean time for the next service completion is calculated as $1/N\mu$, where μ is the service rate of an individual server and N is the number of servers currently working. The mean time for the next service completion is calculated as above by the memoryless property of exponential distribution. See Chapter 2 of Shortle et al.(2018) for the detailed derivation.

Here, we provide the list of basic notations used in the first part of this paper.

- $\lambda(t)$: the incoming rate of customers
- μ_1 : the rate of serving the food
- μ_2 : the rate of eating the food.
- B : the capacity of the tables
- $Q_1(t)$: the length of the queue to wait for the food
- $Q_2(t)$: the length of the queue to find the table
- $p_1(t)$: the proportion of the tables B that is occupied by the customers who are waiting to obtain the food.
- $p_2(t)$: the proportion of the tables B that are actually occupied by customers who are eating.

We compare two policies that can be applied to the same system. In Policy 1, it is not allowed to take the table before one gets the food. In Policy 2, it is allowed to do so. In reality, it can be the case that some of the customers occupy the tables before the food and the others do not. However, assuming diverse possibilities may make the problem complicated to analyze and hence we assume that all the customers uniformly occupy the tables before the food.

Ⅳ. Description of the dynamics

In this section, we provide how the system will evolve given the system parameters. The results of this section will be used in the next section where we derive the stationary state but also it has a potential to be used in further dynamics.

In Policy 1, it is not allowed take the table before the food. Hence, customers will first line up in $Q_1(t)$ after they join the system. Note that $Q_2(t)$ can pile up only when $p_2(t)=1$. If $p_2(t)<1$, i.e., there are remaining seats, customers need not queue up. Hence, the case division for the following proposition is made.

Proposition 1 (Dynamics in Policy 1).

For each of cases divided by the variable quantities, $Q_1(t)$, $Q_2(t)$ and $p_2(t)$ will follow the ordinary differential equations given as follow.

1) $Q_1(t)>0, Q_2(t)>0, p_2(t)=1$:

$$\frac{dQ_1(t)}{dt}=\lambda(t)-\mu_1,$$

$$\frac{dQ_2(t)}{dt}=\mu_1-p_2(t)B\mu_2=\mu_1-B\mu_2,$$

$$\frac{dp_2(t)}{dt}B=0.$$

2) $Q_1(t)>0, Q_2(t)=0, p_2(t)<1$:

$$\frac{dQ_1(t)}{dt}=\lambda(t)-\mu_1,\quad \frac{dQ_2(t)}{dt}=0,$$

$$\frac{dp_2(t)}{dt}B=\mu_1-p_2(t)B\mu_2.$$

3) $Q_1(t)=0, Q_2(t)>0, p_2(t)=1$:

$$\frac{dQ_1(t)}{dt}=\max\{0,\lambda(t)-\mu_1\},$$

$$\frac{dQ_2(t)}{dt}=\min\{\lambda(t),\mu_1\}-p_2(t)B\mu_2$$

$$=\min\{\lambda(t),\mu_1\}-B\mu_2,$$

$$\frac{dp_2(t)}{dt}B=0.$$

4) $Q_1(t)=0, Q_2(t)=0, p_2(t)<1$:

$$\frac{dQ_1(t)}{dt}=\max\{0,\lambda(t)-\mu_1\},$$

$$\frac{dQ_2(t)}{dt}=0,$$

$$\frac{dp_2(t)}{dt}B=\min\{\lambda(t),\mu_1\}-p_2(t)B\mu_2.$$

Proof.

$Q_1(t)$ increase with the rate of $\lambda(t)$ since all the incoming customers first arrives at $Q_1(t)$. If $Q_1(t)$ is positive, $Q_1(t)$ also decreases with the rate of μ_1 since the queue is served in the first server. Hence the results on $\frac{dQ_1(t)}{dt}$ follows on 1) and 2). However, if $Q_1(t)=0$, it will not decrease even though $\lambda(t)<\mu_1$ and the result on $\frac{dQ_1(t)}{dt}$ follows on 3) and 4).

$Q_2(t)$ is 0 and the rate of increase is also 0 when $p_2(t) < 1$, since there is no need to queue up when the tables are available. Hence, $\frac{dQ_2(t)}{dt}$ in 2) and 4) comes. If $p_2(t) = 1$. $Q_2(t)$ increases as $Q_1(t)$ is depleted, i.e., as the food is served. If $Q_1(t) > 0$, that rate is μ_1 and if $Q_1(t) = 0$, the rate is the minimum of $\lambda(t)$ and μ_1; if $\lambda(t)$ is smaller, all the incoming customers immediately pass the food service and go to the second queue and if μ_1 is smaller, only μ_1 rate of customers will conveyed to the second queue and $Q_1(t)$ will increase with the rate of $\lambda(t) - \mu_1$. $Q_2(t)$ decrease as the tables are used up, i.e., $p_2(t)B\mu_2$. If $p_2(t) = 1$, that value is $B\mu_2$ and our results on $\frac{dQ_2(t)}{dt}$ in 1) and 3) are proved.

Lastly, $p_2(t)B$ is the number of tables that are occupied and used for eating. It is kept full if $Q_2(t) > 0$ since the tables are left by the customers who finished eating. It will immediately be replenished by the ones who were waiting in the second queue. Hence, we obtain the results on $\frac{dp_2(t)}{dt}B$ in 1) and 3). If $p_2(t) < 1$ and $Q_2(t) = 0$, the customers who obtained the food immediately take the table. Hence, $p_2(t)B$ increases with the output rate of $Q_1(t)$. Also the customers finish the food in the table with the rate of $p_2(t)B\mu_2$ as was in the assumption. Hence, the results on $\frac{dp_2(t)}{dt}B$ in 2) and 4) follow. □

The dynamics of Policy 2 is more complicated. To avoid additional complexity, we assume all the customers first take the table. Hence, $Q_2(t)$ is encountered before $Q_1(t)$ for all the customers. The following lemma shows properties that should generally hold in Policy 2.

Lemma 1.
In Policy 2, the following equations hold.

$$Q_1(t) = p_1(t)B,$$
$$Q_2(t) = 0 \text{ if } p_1(t) + p_2(t) < 1.$$

Proof.
$Q_1(t) = p_2(t)B$ is given almost from the assumption of the system. In Policy 2, all the customers first take tables, i.e., become part of B, and immediately queue up at $Q_1(t)$. Hence, the portion B that are occupied by pre-eating customers, i.e., $p_2(t)B$ and the queue length of the second queue are equivalent.

If $p_1(t) + p_2(t) < 1$, there are still tables that are not occupied by any kind of customers. Hence, no customers will pile up in the queue for the table and hence $Q_2(t) = 0$ follows. □

Proposition 2 shows equations that should hold in Policy 2. Lemma 1 is used in categorizing the cases.

Proposition 2 (Dynamics in Policy 2).

In the following cases, each set of equations and differential equations hold under Policy 2.

1) $Q_1(t)>0, p_1(t)>0, Q_2(t)>0,$
$p_1(t)+p_2(t)=1$:

$$\frac{dQ_2(t)}{dt}=\lambda(t)-p_2(t)B\mu_2,$$
$$\frac{dQ_1(t)}{dt}=\frac{dp_1(t)}{dt}B=p_2(t)dB\mu_2-\mu_1,$$
$$\frac{dp_2(t)}{dt}B=\mu_1-p_2(t)B\mu_2.$$

2) $Q_1(t)=0, p_1(t)=0, Q_2(t)>0,$
$p_1(t)+p_2(t)=1$, $p_2(t)=1$:

$$\frac{dQ_2(t)}{dt}=\lambda(t)-B\mu_2,$$
$$\frac{dQ_1(t)}{dt}=\frac{dp_1(t)}{dt}B=\max\{0,B\mu_2-\mu_1\},$$
$$\frac{dp_2(t)}{dt}B=\min\{0,\mu_1-B\mu_2\}.$$

3) $Q_1(t)>0, p_1(t)>0, Q_2(t)=0,$
$p_1(t)+p_2(t)<1$:

$$\frac{dQ_2(t)}{dt}=0,\ \frac{dQ_1(t)}{dt}=\frac{dp_1(t)}{dt}B=\lambda(t)-\mu_1,$$
$$\frac{dp_2(t)}{dt}B=\mu_1-p_2(t)B\mu_2.$$

4) $Q_1(t)=0, p_1(t)=0, Q_2(t)=0,$
$p_1(t)+p_2(t)<1$:

$$\frac{dQ_2(t)}{dt}=0,$$
$$\frac{dQ_1(t)}{dt}=\frac{dp_1(t)}{dt}B=\max\{0,\lambda(t)-\mu_1\},$$
$$\frac{dp_2(t)}{dt}B=\min\{\lambda(t),\mu_1\}-p_2(t)B\mu_2.$$

Proof.

In 1), all the queues are positive. We first start with the change of $p_2(t)B$. $p_2(t)B$ decreases with the rate of $p_2(t)B\mu_2$ and increases with the output rate of $Q_1(t)$. Since $Q_1(t)>0$, the food server works without slacks and hence the output rate of $Q_1(t)$ is μ_1 and $\frac{dp_2(t)}{dt}B=\mu_1-p_2(t)B\mu_2$. As the customers go out from $p_2(t)B$, the tables become available and the customers in $Q_2(t)$ obtain them and become a part of $Q_1(t)$. Hence, the input rate of $Q_1(t)$ is $p_2(t)B\mu_2$ and the output rate of $Q_2(t)$ is also $p_2(t)B\mu_2$. Finally, the input rate of $Q_2(t)$ is $\lambda(t)$, since customers first queue up at the second queue. Hence, all the results in 1) follow.

In 2), $p_1(t)=0$ by $Q_1(t)=0$. Also by $Q_2(t)>0$, $p_1(t)+p_2(t)=1$. Hence, $p_2(t)=1$. We start the discussion from $Q_2(t)$. The input rate is $\lambda(t)$ as was in 1). The output rate is the same as the output rate of $p_2(t)B$ as was in 1) and since $p_2(t)=1$, the output rate is $B\mu_2$. Hence, $\frac{dQ_2(t)}{dt}=\lambda(t)-B\mu_2$. If $Q_1(t)>0$ as was in 1), $\frac{dQ_1(t)}{dt}$ would be $B\mu_2-\mu_1$. Since $Q_1(t)=0$ in 2), it cannot decrease and hence $\frac{dQ_1(t)}{dt}=\max\{0,B\mu_2-\mu_1\}$. Lastly, the output rate of $p_2(t)B$ is $B\mu_2$ while the input rate is $\min\{\mu_{1,}$

$B\mu_2\}$. Since $p_2(t)=1$ in 2), it cannot increase. Hence, $\frac{dp_2(t)}{dt}B=\min\{0,\min\{\mu_1,B\mu_2\}-B\mu_2\}=\min\{0,\mu_1-B\mu_2\}$.

In 3), all the incoming customers directly take the tables and queue up at $Q_1(t)$ since $p_1(t)+p_2(t)<1$. Hence, $\frac{dQ_2(t)}{dt}=0$ and $\frac{dQ_1(t)}{dt}=\lambda(t)-\mu_1$. $\frac{dp_2(t)}{dt}B$ follows as in 1). In 4), $\frac{dQ_2(t)}{dt}=0$ as in 3). Since $Q_1(t)=0$, it cannot decrease and hence $\frac{dQ_1(t)}{dt}$ is $\max\{0,\lambda(t)-\mu_1\}$ instead of $\lambda(t)-\mu_1$. Lastly, $\frac{dp_2(t)}{dt}B$ follows as in the previous cases. □

Ⅴ. Stationary state and social welfare analysis

A stationary state in queueing theory is a state where the variable distributions do not change as time passes. Hence, performance of the system, such as average number of people in the system or average waiting time of the customers also does not change. Some of stationary states are also steady states which the system reaches in the long run. Steady states or stationary states are usually the focus of analysis since they are unique and they facilitate mathematical analysis. In this section, we characterize the stationary state of the system and find the social welfare in it. Since we are working with the fluid approximation of the system where all the randomness is vanished, all the variables and performances will be calculated as fixed value rather than expectations of variables with non-trivial distributions.

We are also interested in the welfare analysis of the system. Starting from Naor (1969), there have been two main kinds of objective functions that are considered in the literature - revenue and social welfare. The revenue is the one the service provider earns and the social welfare is the amount of revenue added by the net utility of the customers. In this paper, we focus on the social welfare in the stationary state. In the microeconomic framework for queueing systems, some of the customer behaviors are considered as the result of decision making. In our model, joining decision is modeled as the result of utility maximization. Not every customer who observes the queue enters the system. Customers compare the reward - the positive utility they will enjoy from obtaining the service - and the related cost. Only when the reward is larger, one joins the queue. This framework naturally enables the calculation the social welfare.

We present additional definitions for the social welfare analysis. Since customers have a choice to join or not, there are two kinds of incoming rates - the potential incoming rate

$\Lambda(t)$ and the actual incoming rate $\lambda(t)$. Only the customers who finds the expected utility of entering the system to be positive make an actual entry and hence $\lambda(t) \leq \Lambda(t)$. The reward from the service is assumed to be a continuous random variable R that follows the cumulative distribution function $F(r)$. We assume there exist the lower bound r_L and the upper bound r_U for the support of R, i.e., $F(r_L)=0$ and $F(r_U)=1$. When the customer joins the queue, the waiting cost will incur. We assume that the waiting cost is linear to the waiting time. c_1 is defined as the cost coefficient for the tables and c_2 is defined as the one for the food. Also, we denote the waiting time from Q_i as W_i for $i=1,2$. Hence, if a customer has waited in the queue for the table for 2 units of time ($W_1=2$) and in the queue for the food for 3 units of time ($W_2=3$), the total cost the customer experience is $2c_1+3c_2$.

A customer's net utility is given as follows:

$$R-c_1E[W_1]-c_2E[W_2]$$

Once customers observe the system, they calculate the above net utility and observe whether it is positive or not. Since the net utility is an increasing function of the reward, we can find the threshold, say $\bar{r}(t)$, such that the customers with $R<\bar{r}(t)$ leave the system without obtaining the service and the customers with $R \geq \bar{r}(t)$ enter.

The following shows the list of the additional notations that enable the social welfare analysis.

- $\Lambda(t)$: the potential incoming rate.
- W_1 : the waiting time experienced in the queue for the table
- W_2 : the waiting time experienced in the queue for the food.
- c_1 : the cost related to the queue for the table.
- c_2 : the cost related to the queue for the food.
- R : the reward from the service. It is a continuous random variable distributed with the cumulative distribution $F(r)$.
- r_L : the lower bound for the support of R, i.e., $F(r_L)=0$.
- r_U : the upper bound for the support of R, i.e., $F(r_H)=1$.
- $\bar{r}(t)$: the threshold of the reward.

We will find the stationary states of the system. The notations that evolve with the time have been denoted with t until now. For example, the incoming rate is given as $\lambda(t)$. The stationary versions of them will be denoted with ∞. Hence, $\lambda(\infty)$ will denote the stationary incoming rate, if it exists. Also, we will now assume that the potential incoming rate is stationary to facilitate the discussion in the stationary state.

Proposition 3 shows the parametric conditions that allow stationary states and properties of stationary queue lengths for Policy 1.

Proposition 3 (Stationary states in Policy 1).

Depending on the conditions on $\lambda(\infty)$, μ_1 and $B\mu_2$, the stationary states $Q_1(\infty)$, $Q_2(\infty)$, $p_1(\infty)$ and $p_2(\infty)$ satisfy the followings.

1) $\lambda(\infty) = \mu_1 = B\mu_2$:
$Q_1(\infty) > 0, Q_2(\infty) > 0, p_1(\infty) = 0.$

2) $\lambda(\infty) = \mu_1 < B\mu_2$:
$Q_1(\infty) > 0, Q_2(\infty) = 0, p_2(\infty) < 1,$
$\mu_1 = p_2(\infty)B\mu_2.$

3) $\lambda(\infty) = B\mu_2 < \mu_1$:
$Q_1(\infty) = 0, Q_2(\infty) > 0, p_2(\infty) = 1.$

4) $\lambda(\infty) < \min\{\mu_1, B\mu_2\}$:
$Q_1(\infty) = 0, Q_2(\infty) = 0, p_2(\infty) < 1.$

Proof.

We start with Proposition 1 which depicted the differential equations that should be satisfied. In the stationary states, those differential equations should take the value of 0 and it gives the condition in the current theorem.

To make $Q_1(\infty) > 0, Q_2(\infty) > 0, p_2(\infty) = 1$, the followings should hold. See 1) of Proposition 1.

$$0 = \frac{dQ_1(t)}{dt} = \lambda(\infty) - \mu_1,$$
$$0 = \frac{dQ_2(t)}{dt} = \mu_1 - p_2(\infty)B\mu_2 = \mu_1 - B\mu_2,$$
$$\frac{dp_2(t)}{dt}B = 0.$$

Rearranging $\lambda(\infty) = \mu_1 = B\mu_2$ should hold and it makes 1).

If $Q_1(\infty) > 0, Q_2(\infty) = 0, p_2(\infty) < 1$, the following should hold. See 2) of Proposition 1.

$$0 = \frac{dQ_1(t)}{dt} = \lambda(\infty) - \mu_1,\ \frac{dQ_2(t)}{dt} = 0,$$
$$0 = \frac{dp_2(t)}{dt}B = \mu_1 - p_2(\infty)B\mu_2,$$

Rearranging we have $\lambda(\infty) = \mu_1 = p_2(\infty)B\mu_2$ and since $p_2(t) < 1$, $\mu_1 < B\mu_2$. Hence, 2) follows.

To make $Q_1(\infty) = 0, Q_2(\infty) > 0, p_2(\infty) = 1$, the following should hold by 3) of Proposition 1.

$$0 = \frac{dQ_1(t)}{dt} = \max\{0, \lambda(\infty) - \mu_1\},$$
$$0 = \frac{dQ_2(t)}{dt} = \min\{\lambda(\infty), \mu_1\} - B\mu_2,$$
$$0 = \frac{dp_2(t)}{dt}B = 0.$$

Rearranging, $\min\{\lambda(\infty), \mu_1\} = B\mu_2$ and $\lambda(\infty) < \mu_1$. Hence, $\lambda(\infty) = B\mu_2 < \mu_1$ and 3) follows.

To make $Q_1(\infty) = 0, Q_2(\infty) = 0, p_2(\infty) < 1$, the following should hold by 4) of Proposition 1.

$$0=\frac{dQ_1(t)}{dt}=\max\{0,\lambda(\infty)-\mu_1\},\frac{dQ_2(t)}{dt}=0,$$
$$0=\frac{dp_2(t)}{dt}=\min\{\lambda(\infty),\mu_1\}-p_2(\infty)B\mu_2.$$

Rearranging, $\lambda(\infty)<\mu_1$ and $\lambda(\infty)-p_2(\infty)B\mu_2=0$. Hence, $\lambda(\infty)<\min\{\mu_1,B\mu_2\}$ and 4) follows. □

Proposition 4 shows the parametric conditions that allow stationary states and properties of stationary queue lengths for Policy 2.

Proposition 4 (Stationary states in Policy 2).

For each of the cases, the followings hold.

1) $\lambda(\infty)=\mu_1\le B\mu_2$:
   $\mu_1=p_1(\infty)B\mu_2$,
   Either $Q_1(\infty)<B-\frac{\mu_1}{\mu_2}$ and $Q_2(\infty)=0$ or $Q_1(\infty)=B-\frac{\mu_1}{\mu_2}$ and $Q_2(\infty)>0$.

2) $\lambda(\infty)=B\mu_2<\mu_1$:
   $Q_1(\infty)=0(p_1(\infty)=0)$,
   $Q_2(\infty)>0(p_1(\infty)+p_2(\infty)=1)$,
   $p_2(\infty)=1$.

3) $\lambda(\infty)<\min\{\mu_1,B\mu_2\}$:
   $Q_1(\infty)=0(p_1(t)=0)$,
   $Q_2(\infty)=0(p_1(\infty)+p_2(\infty)<1)$.

Proof.

Note that $p_1(\infty)+p_2(\infty)\le 1$ should always hold by the assumption. $p_1(\infty)+p_2(\infty)\le 1$ can be equivalently depicted as $\frac{\mu_1}{B\mu_2}+\frac{Q_1(\infty)}{B}\le 1$ since $Q_1(\infty)=p_1(\infty)B$ and $\mu_1=p_2(\infty)B\mu_2$. If $p_1(\infty)+p_2(\infty)=1$, $\frac{\mu_1}{B\mu_2}+\frac{Q_1(\infty)}{B}=1$ and $Q_1(\infty)=B\left(1-\frac{\mu_1}{B\mu_2}\right)=B-\frac{\mu_1}{\mu_2}$.

To make $Q_1(\infty)>0, p_1(\infty)>0, Q_2(\infty)>0, p_1(\infty)+p_2(\infty)=1$, the following should hold by 1) of Proposition 2.

$$0=\frac{dQ_2(t)}{dt}=\lambda(\infty)-p_2(\infty)B\mu_2,$$
$$0=\frac{dQ_1(t)}{dt}=\frac{dp_1(t)}{dt}B=p_2(\infty)B\mu_2-\mu_1,$$
$$0=\frac{dp_2(t)}{dt}B=\mu_1-p_2(\infty)B\mu_2.$$

Rearranging we have $\mu_1=p_2(\infty)B\mu_2=\lambda(\infty)$. Since $p_2(\infty)\le 1$, $\mu_1\le B\mu_2$. By $p_1(\infty)+p_2(\infty)=1$, $Q_1(\infty)=B\left(1-\frac{\mu_1}{B\mu_2}\right)=B-\frac{\mu_1}{\mu_2}$.

To make $Q_1(\infty)>0, p_1(\infty)>0, Q_2(\infty)=0, p_1(\infty)+p_2(\infty)<1$, the following should hold by 3) of Proposition 2.

$$\frac{dQ_2(t)}{dt}=0,$$
$$0=\frac{dQ_1(t)}{dt}=\frac{dp_1(t)}{dt}B=\lambda(\infty)-\mu_1,$$
$$0=\frac{dp_2(t)}{dt}B=\mu_1-p_2(\infty)B\mu_2$$

Rearranging $\lambda(\infty)=\mu_1=p_2(\infty)B\mu_2$. Since $p_2(\infty)\le 1$, $\mu_1 \le B\mu_2$. Since $p_1(\infty)+p_2(\infty)<1$, $Q_1(\infty)<B-\frac{\mu_1}{\mu_2}$ and hence 1) of the current proposition is proved.

To make $Q_1(\infty)=0, p_1(\infty)=0, Q_2(\infty)>0, p_1(\infty)+p_2(\infty)=1$, $p_2(\infty)=1$, the following holds by 2) of Proposition 2.

$$0=\frac{dQ_2(t)}{dt}=\lambda(\infty)-B\mu_2,$$
$$0=\frac{dQ_1(t)}{dt}=\frac{dp_1(t)}{dt}B=\max\{0,B\mu_2-\mu_1\},$$
$$0=\frac{dp_2(t)}{dt}B=\min\{0,\mu_1-B\mu_2\}$$

Rearranging, we have $\lambda(\infty)=B\mu_2$ and $\mu_1 \ge B\mu_2$ and 2) of the current proposition is proved.

To make $Q_1(t)=0, p_1(t)=0, Q_2(t)=0, p_1(t)+p_2(t)<1$, the followings should hold by 4) of Proposition 2.

$$\frac{dQ_2(t)}{dt}=0,$$
$$0=\frac{dQ_1(t)}{dt}=\frac{dp_1(t)}{dt}B=\max\{0,\lambda(\infty)-\mu_1\},$$
$$0=\frac{dp_2(t)}{dt}B=\min\{\lambda(\infty),\mu_1\}-p_2(\infty)B\mu_2$$

Rearranging we have $\lambda(\infty)\le\mu_1$ and $\lambda(\infty)=p_2(\infty)B\mu_2$. Hence, $\lambda(\infty)<\min\{\lambda(\infty), B\mu_2\}$ and 3) of the current proposition is proved. □

Finally, we compare the utilities generated from each of the policies. The system utility is given as the reward from the service minus the waiting cost for the whole customers as follows. Since the utility of the customers who do not choose to enter the system is normalized to 0, we only need to sum the net utility of the customers who actually enter.

(Total reward) – (Total waiting cost from Q_1) – (Total waiting cost from Q_2)

Since we are dealing with the fluid models, the summation of the net utilities for the whole customers is represented as an integral. The following theorem is the main result of this paper. The result shows that Policy 1 and Policy 2 show the same system utility performance in the stationary states.

Theorem 1 (Identity of the system utilities). The system utility is not dependent on the policies.

Proof.

We will exclude the case $\mu_1=B\mu_2$ which occurs with probability 0 and hence first, assume that $\mu_1<B\mu_2$ and $\Lambda(\infty)>\mu_1$. By 2) of Proposition 3, $Q_1(\infty)>0, Q_2(\infty)=0, p_2(\infty)<1$ makes a stationary state with $\lambda(\infty)=\mu_1=p_2(\infty)B\mu_2$ in Policy 1. Then, there exists $\bar{r}$ that satisfies the following.

$$\lambda(\infty)=\Lambda(\infty)\big(F(r_U)-F(\bar{r})\big)$$
$$=\Lambda(\infty)\big(1-F(\bar{r})\big)=\mu_1$$

$\bar{r}$ can be calculated as follows:

$$\bar{r}=F^{-1}\left(1-\frac{\mu_1}{\Lambda(\infty)}\right)$$

The accumulation of queues occurs until the customers with the reward $\bar{r}$ be indifferent between entering the service or renege. In case of $\mu_1<B\mu_2$, only Q_1 accumulates. Since the we are working with fluid limits, the waiting time is deterministically given as the queue length divided by the service rate. Therefore, $\bar{r}-c_1W_1(\infty)=\bar{r}-c_1\frac{Q_1(\infty)}{\mu_1}=0$ holds and hence $Q_1(\infty)=\frac{\mu_1\bar{r}}{c_1}$.

The system reward in a unit time is given as $\Lambda(\infty)\int_{\bar{r}}^{r_U}rf(r)dr$. Among the whole potential customers, $f(r)dr$ portion of customers will have the reward of r and using the same way the expectations of continuous variables are calculated the system reward is given as above with $\bar{r}=F^{-1}\left(1-\frac{\mu_1}{\Lambda(\infty)}\right)$.

By the linearity of the waiting time cost function, the waiting cost in a unit time is given as $c_1Q_1(\infty)$, the system utility is calculated as

$$\Lambda(\infty)\int_{\bar{r}}^{r_U}rf(r)dr-c_1Q_1(\infty)$$
$$=\Lambda(\infty)\int_{\bar{r}}^{r_U}rf(r)dr-c_1\frac{\mu_1\bar{r}}{c_1}$$
$$=\Lambda(\infty)\int_{\bar{r}}^{r_U}rf(r)dr-\mu_1\bar{r}.$$

Now we compare the above with the system utility from Policy 2. By 2) of Proposition 3, $\lambda(\infty)=\mu_1=p_1(\infty)B\mu_2$ holds and either $Q_1(\infty)=B-\frac{\mu_1}{\mu_2}$ and $Q_2(\infty)>0$ or $Q_1(\infty)<B-\frac{\mu_1}{\mu_2}$ and $Q_2(\infty)=0$ holds. By $\lambda(\infty)=\mu_1$ we can calculate the threshold reward again as follows.

$$\lambda(\infty)=\Lambda(\infty)\big(F(r_U)-F(\bar{r})\big)$$
$$=\Lambda(\infty)\big(1-F(\bar{r})\big)=\mu_1.$$

Hence, the threshold $\bar{r}$ is the same as in Policy 1, i.e., $\bar{r}=F^{-1}\left(1-\frac{\mu_1}{\Lambda(\infty)}\right)$.

Depending on the system parameters, the $Q_2(\infty)$ may be positive or zero. First, we assume $Q_2(\infty)>0$ and hence $Q_1(\infty)=B-\frac{\mu_1}{\mu_2}$ and see whether there is a contradiction. Since the customer with the threshold reward $\bar{r}$ will be indifferent between entering the service and not, the following holds.

$$\bar{r}-c_1\frac{Q_1(\infty)}{\mu_1}-c_2\frac{Q_2(\infty)}{\mu_1}=0.$$

Then, the stationary $Q_2(\infty)$ is calculated as

$$Q_2(\infty)=\frac{\mu_1\bar{r}-c_1B+c_1\frac{\mu_1}{\mu_2}}{c_2}.$$

If the above value is positive we are in the case with $Q_2(\infty)>0$. If it is not the case $Q_2(\infty)=0$ and $Q_1(\infty)<B-\frac{\mu_1}{\mu_2}$. In case of $Q_2(\infty)>0$, the system utility is calculated as follows:

$$\begin{aligned}
&\Lambda(\infty)\int_{\bar{r}}^{r_U} rf(r)dr-c_1\left(B-\frac{\mu_1}{\mu_2}\right)-c_2\frac{\mu_1\bar{r}-c_1B+c_1\frac{\mu_1}{\mu_2}}{c_2}\\
&=\Lambda(\infty)\int_{\bar{r}}^{r_U} rf(r)dr-c_1\left(B-\frac{\mu_1}{\mu_2}\right)-\left(\mu_1\bar{r}-c_1B+c_1\frac{\mu_1}{\mu_2}\right)\\
&=\Lambda(\infty)\int_{\bar{r}}^{r_U} rf(r)dr-c_1B+c_1\frac{\mu_1}{\mu_2}-\mu_1\bar{r}+c_1B-c_1\frac{\mu_1}{\mu_2}\\
&=\Lambda(\infty)\int_{\bar{r}}^{r_U} rf(r)dr+c_1\frac{\mu_1}{\mu_2}-\mu_1\bar{r}-c_1\frac{\mu_1}{\mu_2}\\
&=\Lambda(\infty)\int_{\bar{r}}^{r_U} rf(r)dr-\mu_1\bar{r}.
\end{aligned}$$

The system utility formula is the same as in Policy 1. Since $\bar{r}$ is also the same, we have the equivalence in the case of $Q_2(\infty)>0$.

Now suppose $Q_2(\infty)=0$ and $Q_1(\infty)<B-\frac{\mu_1}{\mu_2}$. Since the threshold customer will have the zero utility as before, $\bar{r}-c_1\frac{Q_1(\infty)}{\mu_1}=0$ holds.

Since $Q_2(\infty)=0$, the system utility is calculated as

$$\begin{aligned}
&\Lambda(\infty)\int_{\bar{r}}^{r_U} rf(r)dr-c_1Q_1(\infty)\\
&=\Lambda(\infty)\int_{\bar{r}}^{r_U} rf(r)dr-\mu_1\bar{r}
\end{aligned}$$

with $\bar{r}=F^{-1}\left(1-\frac{\mu_1}{\Lambda(\infty)}\right)$ again and the equivalence still holds in this case.

Now assume $\mu_1>B\mu_2$ and $\Lambda(\infty)>B\mu_2$. By 3) of Proposition 3, $\lambda(\infty)=B\mu_2<\mu_1$ $Q_1(\infty)=0$, $Q_2(\infty)>0, p_2(\infty)=1$ holds in stationarity when Policy 1 is adopted. Now the threshold is given from the following equation:

$$\begin{aligned}
\lambda(\infty)&=\Lambda(\infty)\big(F(r_U)-F(\bar{r})\big)\\
&=\Lambda(\infty)\big(1-F(\bar{r})\big)=B\mu_2.
\end{aligned}$$

Hence, $\bar{r}=F^{-1}\left(1-\frac{B\mu_2}{\Lambda(\infty)}\right)$.

The queue accumulates until customers with $\bar{r}$ have the utility of 0. Since only Q_2 accumulates, $\bar{r}-c_2\frac{Q_2(\infty)}{B\mu_2}=0$ and hence $Q_2(\infty)=\frac{B\mu_2\bar{r}}{c_2}$.

The system utility in Policy 1 is then,

$$\begin{aligned}
&\Lambda(\infty)\int_{\bar{r}}^{r_U} rf(r)dr-c_2Q_2(\infty)\\
&=\Lambda(\infty)\int_{\bar{r}}^{r_U} rf(r)dr-B\mu_2\bar{r}.
\end{aligned}$$

When Policy 2 is applied, $\lambda(\infty)=B\mu_2<\mu_1$,

$Q_1(\infty)=0(p_1(\infty)=0)$, $Q_2(\infty)>0(p_1(\infty)+p_2(\infty)=1)$, $p_2(\infty)=1$. Hence, by the same reasoning as in Policy 1, we obtained the same system utility with the same $\bar{r}$ and hence the equivalence is also proved when $\mu_1 > B\mu_2$.

Finally in case of $\Lambda(\infty)<\min\{\mu_1,\ B\mu_2\}$, whether in Policy 1 or 2, all the potentjial customers can enter without incurring any waiting cost since in $Q_1(\infty)=Q_2(\infty)=0$

Everyone will enter without incurring waiting cost and hence also the equivalence holds. □

Ⅵ. Conclusion

We have analyzed the problem of cafeteria queues where we have two choices of polices regarding the sequence the customer may go through – first taking the table and obtain the food and first obtain the food and find the table. Since executing an exact analysis is possibly a daunting job, we have used fluid limits to facilitate calculation of the performances. In the fluid limits, the effect from the first order moment of the variables survives while variability, especially the second order moment effect vanishes. We have also assumed that if the table taking before the food is allowed, all the customers will uniformly do it first.

First, the dynamics of the queue lengths for the table and the food are derived as with ordinary differential equation in different parametric assumptions. We use these general results to find the property of stationary states. Finally, the social welfare in the stationary state incurred from each of the policies are calculated and it turned out that the social welfares are the same regardless of the policies.

Hence, we can conclude that the choice of the policy does not affect the social welfare in the stationary state at least in our model where the first order effect is important. We can infer that the notion that taking the table first is socially sub-optimal is not well-supported analytically especially under the assumption that none of the customers first wait for the table and take the food. Service providers can take managerial implications from this point – whatever policy is employed, the social welfare is the same if it is uniformly applied to all the customer. The conditional part is especially important and it can explain why we are sometimes asked by service providers when ordering the food whether we had already reserved the seat. It can be interpreted as urging customers to behave uniformly. Our work is the first one to raise the problem of whether or not to allow preoccupation of seats and hence is distinct from existing literature on cafeteria queues, e.g., June & Jain(2010), Weber & Weiss(1994) and Füßler et al.(2019).

However, there are several aspects that are not dealt in this paper that may change our conclusion on the identity of social welfare, Effectively addressing these will make good extension from our framework. The most important extension will be on adding customer heterogeneity in their behavior. Even though taking the table first is allowed, some of the customers do not or cannot take it. Then, it is possible that inequality between the two groups – customers who take the table first and the one who do not – may occur. Since in the policy where preoccupation of the table is not allowed, inequality among customers will not appear, this difference may support the notion that disallowance of preoccupation is desirable. Similarly, the customers may differ in the size of the group they belong to. Customers came alone will not have ways to reserve the seat, while ones in groups can utilize division of labor and let one member to reserve and others to take foods. This heterogeneity will also increase the inequality of the system. A fact that real-world decision maker does not always make optimal solutions (Kwak, 2014) can also be incorporated in the model.

Additionally, using a different analytic modeling may change the conclusion. There are two main aspects in our modeling – using fluid limit and stationary states. Using an exact analysis or approximation with Brownian motion where the second order moment is not vanished may convert our qualitative results. If the second moment is incorporated in the model by Brownian approximation, we can develop a criterion that is based on not only the expectation but also on variance. Then, usual preference toward the policy that does not allow preoccupation can be justified by smaller variance in the waiting times. Also comparing performance index for transient states rather than the stationary states will enrich our understanding of cafeteria queues. For example, it can be the case that one policy outperform the other in clearing times when there are given number of customers and if the customers do not have a choice to renege the system. The basic framework that we offer in this paper may work as a base for those possibly fruitful extensions. Our contribution lies in first raising the policy selection problem in cafeteria queues and also providing a stylized model to study it.

References

Afeche, P., M. Araghi and O. Baron(2017), "Customer Acquisition, Retention, and Service Access Quality: Optimal Advertising, Capacity Level, and Capacity allocation," *Manufacturing & Service Operations Management*, 19(4), pp. 674-691

Allon, G., A. Bassamboo and E. B. Cil(2017), "Skill Management in Large-Scale Service Marketplaces," *Production and Operations Management*, 26(11), pp.2050-2070.

Allon, G. and I. Gurvich(2010), "Pricing and Dimensioning Competing Large-Scale Service Providers," *Manufacturing & Service Operations*

Management, 12(3), pp.449-469.
Ata, B., A. Skaro and S. Tayur(2017), "OrganJet: Overcoming Geographical Disparities in Access to Deceased Donor Kidneys in the United States," *Management Science*, 63(9), pp.2776-2794.
Cho, J. and S. Kim(2007), "Service Waiting: How Wait Times Affect Service Evaluations," *Korean Management Review*, 36(7), pp.1785-1810.
Economou, A. and A. Manou(2016), "Strategic Behavior in an Observable Fluid Queue with an Alternating Service Process," *European Journal of Operational Research*, 254(1), pp.148-160.
Füßler, D., S. Fedtke and N. Boysen(2019), "The Cafeteria Problem: Order Sequencing and Picker Routing in On-the-Line Picking Systems," *OR Spectrum*, 41, pp.727-756.
Hassin, R.(2016), *Rational Queueing*, CRC press, Florida.
Hassin, R. and M. Haviv(2003), *To Queue or Not to Queue: Equilibrium Behavior in Queueing Systems (Vol. 59)*, Springer Science & Business Media, Germany.
Juneja, S. and R. Jain(2010), "The Concert/Cafeteria Queueing Problem: A Game of Arrivals," In 4th International ICST Conference on Performance Evaluation Methodologies and Tools, Pisa, Italy, 2009.
Kwak, J.(2014), "Application of Newsvendor Experiments to Education of Operations Management," *Korean Management Review*, 43(5), pp.1519-1528.
Lee, K. W. and C. U. Lambert(2006), "Using Simulation to Manage Waiting Time in a Cafeteria," *Information Technology in Hospitality*, 4(4), pp.127-141.
Liu, J., X. Xu, S. Wang and D. Yue(2021), "Equilibrium Analysis of the Fluid Model with Two Types of Parallel Customers and Breakdown," *Communications in Statistics-Theory and Methods*, 50(24), pp.5792-5805.
Logothetis, D., A. Manou and A. Economou(2022), "The Impact of Reneging on a Fluid On-Off Queue with Strategic Customers," *Annals of Operations Research*.
Maglaras, C. and A. Zeevi, (2005), "The Impact of Reneging on a Fluid On-Off Queue with Strategic Customers," *Operations Research*, 53(2), pp.242-262.
Naor, P.(1969), "The Regulation of Queue Size by Levying Tolls," *Econometrica: Journal of the Econometric Society*, 37, pp.15-24.
Park, M., B. Park and K. Park(1999), "An Analysis on the Ship Handling System at a Containter Terminal using Queueing Theory and Simulation Simultaneously," *Korean Management Review*, 28(1), pp.151-166.
Rajagopal, S., V. G. Kulkarni and S. Stidham(1995), "Optimal Flow Control of a Stochastic Fluid-Flow System," *IEEE Journal on Selected Areas in Communications*, 13(7), pp.1219-1228.
Saini, A. N.(2017), "Is It OK to Save Seats in a Crowded Restaurant?" https://www.sporkful.com/is-it-ok-to-save-seats-in-a-crowded-restaurant/(retrieved August 2023).
Shortle, J. F., J. M. Thompson, D. Gross and C. M. Harris(2018), *Fundamentals of Queueing Theory (Vol. 399),* John Wiley & Sons, New Jersey.
Stout Jr, W. A.(1995), "Modeling a Hospital Main Cafeteria," In *Proceedings of the 27th conference on Winter simulation,* Arlington,

Virginia.
Wang, S. and X. Xu(2021), "Equilibrium Strategies of the Fluid Queue with Working Vacation," *Operational Research*, 21, pp.1211-1228.
Weber, R. R. and G. Weiss(1994), "The Cafeteria Process－Tandem Queues with 0-1 Dependent Service Times and the Bowl Shape Phenomenon," *Operations Research*, 42(5), pp.895-912.

• The author Seung Bum Soh is an Assistant Professor of Management Science at Yonsei Business School. He graduated from the College of Business at Seoul National University, where he earned a bachelor's and a master's degree. He obtained his PhD degree from the Kellogg School of Management at Northwestern University. He served as a Postdoctoral Researcher in the department of Industrial Engineering at Hong Kong University of Science and Technology and as an Assistant Professor at Sejong University. His main research interests include queueing theory, recommendation systems and portfolio management.