TEL-AVIV UNIVERSITY
RAYMOND AND BEVERLY SACKLER
FACULTY OF EXACT SCIENCES
SCHOOL OF MATHEMATICAL SCIENCES,
DEPARTMENT OF STATISTICS AND OPERATION RESEARCH

Equilibrium solutions in the observable $M/M/1$ queue with overtaking

Thesis submitted in partial fulfillment of
requirments for the M.Sc. degree in the faculty of Exact Sciences,
school of Mathematical Sciences, Tel-Aviv University

by Jenny Erlichman

The research work for this thesis has been carried out at Tel-Aviv University under the supervision of Prof. Refael Hassin

September 23, 2009

I would like to express my profound gratitude to Professor Refael Hassin for all his guidance, help and support.

Abstract

The subject of this paper is a mechanism that allows customers to overtake others. In our system, customers observe the queue length upon arrival, and have the option of overtaking some or all of the customers already present in the queue. Overtaking is associated with a fixed price per overtaken customer. If a customer chooses to overtake some but not all of the present customers, he overtakes the last customers in the queue. Customers incur a fixed cost per every unit of time in the system, and their goal is to minimize their own expected total cost. We would like to characterize the symmetric equilibrium strategies of our model. However, it turns out that this mission is much harder in our system than in related priority queueing systems analyzed in the literature. We consider several types of symmetric strategies and find out that the set of equilibrium symmetric strategies is quite reach and includes surprisingly odd strategies. In addition, we compare overtaking with preemptive priority systems. We assume that the server can induce the customers to choose, among the equilibrium strategies, the one which maximizes its profits. Under this assumption, we compare the server's profits in the two models and find, somewhat surprisingly, that the system of overtaking gives the server higher profits.

Contents

1 Introduction

Priority sale in queueing systems is a common mechanism used to improve service and increase profits. In such regimes, a customer has the option of purchasing priority, out of a menu of options, and obtain service before some others who arrived earlier. Of course, later arrivals who purchase higher priority may overtake the previously overtaking customer, and this may serve as a further incentive to purchase priority. It is expected that customers take all this into consideration when choosing their purchase strategy. Since a customer's strategy responds to other customers strategy, the result is an equilibrium strategic behavior.

The subject of this paper is a different mechanism that allows customers to overtake others. In our system, customers observe the queue length upon arrival, and have the option of overtaking some or all of the customers already present in the queue. As more customers overtake others, the more inclined an individual should be to do so himself. In other words, this is a follow the crowd (FTC) situation. Overtaking is associated with a fixed price per overtaken customer. If a customer chooses to overtake some but not all of the present customers, overtaking applies to the last customers in the queue. Customers incur a fixed cost per every unit of sojourn time in the system, and their goal is to minimize their own expected total cost. Our model assumes a single server Markovian queue where customers are identical in all (statistical) parameters.

We exclude balking, and hence there is no question of social optimality here. However, different strategies affect the rate of profit to the server.

We set up two main goals:

- We would like to characterize the symmetric equilibrium strategies of our model. However, it turns out that this mission is much harder in our system than in related priority queueing systems analyzed in the literature. We consider several types of symmetric strategies and find out that the set of equilibrium symmetric strategies is quite rich and includes surprisingly odd strategies. We characterize some particular families of equilibrium strategies, but it is clear that these are not the only equilibrium strategies. We found some surprising results of unexpected strategies which, for some parameters, are equilibrium strategies. For example, a strategy like: overtaking a single customer when observing one customer in the system upon arrival, overtaking none when observing two or three, overtaking four customers when observing four, overtaking three customers when observing five, and not overtaking any customer otherwise, can be an equilibrium.

- We compare overtaking with preemptive priority systems. We assume that the server can induce the customers to choose, among the equilibrium strategies, the one which maximizes its profits. Under this assumption, we compare the server's profits in the two models and find, somewhat surprisingly, that the system of overtaking gives the server higher profits.

One of the interesting findings in our work is that sometimes it is worthwhile to observe a longer queue since then the customers's expected cost is lower.

The paper is organized as follows: In Section 2 we formally present our model. In Section 3 we survey relevant literature on strategic behavior in queueing systems with priorities. In Section 4 we try to characterize the equilibrium strategies of our model. We find that it is difficult to characterize all strategies that can be equilibrium and therefore consider some attractive special cases. In Section 5 we consider strategies of overtaking k customers if there are at least k customers in the system, and overtaking all of them otherwise. In Subsection 5.1 we ask whether it is worthwhile to observe a longer queue when all customers follow this strategy and the answer is sometimes yes. In Subsections 5.2 and 5.3 we consider mixed strategies. In Subsection 5.2 we consider mixtures between overtaking k or $k-1$ customers, and in Subsection 5.3 we consider mixtures with overtaking k, $k-1$ or $k-2$ customers. In both cases we give necessary and sufficient conditions for these strategies to define an equilibrium. In Section 6 we compare the server's maximum expected profit per customer under equilibrium conditions in two models. The first is our model, and the second is the model analyzed by Adiri and Yechiali [1] and by Hassin and Haviv [9], in which there are two priority classes. In this model two FCFS queues are formed in front of a single server, one for priority customer and the other for ordinary customers. An arriving customer buys priority if and only if the total number of customers in the system is at least a threshold n. There may be numerous equilibria in both models. We assume that the server can choose the equilibrium which maximizes its expected profit. Under this assumption we prove that the server's expected profit per arrival in our model is greater than the server's expected profit per arrival in the second model. Our proof is partially analytic and partially based on a numerical computations. In Section 7 we analyze equilibria where at most one customer is overtaken. We consider pure and mixed threshold strategies. We give necessary and sufficient conditions for these strategies to define an equilibrium. Finally, Section 8 contains concluding remarks and open problems.

2 Model description

In our observable $M/M/1$ model, customers purchase priority. This priority enables overtaking present customers. Upon arrival, a new customer observes the queue length and announces the number of customers that he overtakes. There is a fixed cost C_o per overtaken customer, and customers have homogeneous waiting costs. We assume that the reward the client receives after he is served is high enough so that, he leaves the system (renege) before he is served only when the queue length is very long, but this has a very low probability. Therefore, we assume that there is no reneging. In addition, there is no balking, and a customer cannot overtake after joining the queue. The service discipline is preemptive resume. Let C_w denote the cost per unit of time to a customer for staying in the system (either waiting or being served). All customers have the same waiting time value C_w. We denote the rate of arrival by λ, and the service rate by μ. Note that the case $C_o < \frac{C_w}{\mu}$ has a trivial unique equilibrium since overtaking all present customers is clearly a dominant strategy. Therefore, we assume $\frac{C_w}{\mu} < C_o$.

As more customers overtake others, the more inclined an individual should be to do so himself. In other words, this is a follow the crowd (FTC) situation. The rationale behind this terminology is that in an FTC case, the higher the values selected by the others, the higher is one's best response.

In a Nash equilibrium no customer has anything to gain by changing his or her own strategy unilaterally. In a symmetric equilibrium all customers use the same strategy.

Consider a static version of the model, where the number of customers to be served is fixed and they are all present at the time the service begins. Moreover, no future arrivals are expected. In this case there is a unique equilibrium in which no customer overtakes any other customer. To see why this is true note that from the assumption of our model that $\frac{C_w}{\mu} < C_o$, a dominant best response of the *last customer* is not overtaking, therefore a best response of the customer whose position before the last one is not overtaking either, and if we continue this way the result is that no one is overtaking. In the sequel we show that while never overtaking is always an equilibrium strategy, in the dynamic model there are numerous of the equilibrium strategies as well.

We analyze pure and mixed threshold strategies, and also other strategies in the dynamic model. We look for equilibrium conditions in each of these strategies.

We compare overtaking with preemptive priority systems. We assume that the server can induce the customers to choose among the equilibrium strategies, the one which maximizes its profits. Hence, server's goal is to induce all customers to overtake all customers who are currently in the system.

3 Related literature

There is an extensive literature on priority queues. In this section we review the literature on strategic behavior in queueing systems with priorities and server's profit maximizing under different priority regimes.

There are three basic priority disciplines [11, 3]: preemptive resume, preemptive repeat and non preemptive. In a preemptive resume discipline the service of a customer is interrupted when a customer belonging to a higher priority class arrives, and will be resumed from the point of interruption. In a preemptive repeat discipline the service of a customer is interrupted upon arrival of customer belonging to a higher priority class, and will start from the beginning. In a non preemptive discipline the service of the customer in service is completed even if a customer of higher priority arrives. In our model the priority discipline is preemptive resume.

- Adiri and Yechiali [1] analyzed an M/M/1 model with two priority classes. In their model the priority discipline is preemptive resume and two FCFS queues are formed in front of a single server, one line for priority customers and one for ordinary customers. Upon arrival and after observing the length of the two queues, a customer decides whether to purchase priority. Customers cannot purchase priority while waiting. Adiri and Yechiali assumed a given price for priority and computed an equilibrium solution of the following type: buy priority if and only if the number of customers in the ordinary queue is at least a threshold n. We compare the server's expected profit in this model to the server's expected profit in our model.

  Hassin and Haviv [9] continued this line of research. They extended the set of possible strategies to include mixed strategies of the following kind: for some nonnegative real number $x = n+p$, where n is an integer and $0 \leqslant p < 1$, a customer who observes a total of k customers in the system joins the ordinary queue if $k \leqslant n-1$, does so with probability p if $k = n$, and otherwise buys priority. They show that multiplicity of equilibria is possible. Moreover, in general there is an interval of integer thresholds that define stable equilibria and between each pair of such equilibria there is an unstable mixed equilibrium.

- Kleinrock [12] considers that relative position in queue is determined according to the size of a customers bribe (which is paid before the customer sees the queue length). Such a policy allows the customer himself to affect his own queue position, rather than the classical approach of assuming that a customer is preassigned to some (possibly continuous) priority class. For the case of Poisson arrivals, arbitrary service time distribution, and arbitrary distribution of customer bribe, obtained the average waiting time for customers as a function of their bribe. Both preemptive and nonpreemptive disciplines are considered. Examples are presented for various bribing distributions, which demonstrate that many well-known priority queuing systems are special cases of this bribing situation. Furthermore, a cost function is defined after introducing the notion of an impatience factor (which converts seconds of wait into dollars). Conditions for optimum bribing are then determined, where the optimization refers to minimizing the average cost subject to a mean bribe constraint. An example for exponential service and exponential bribing is carried out and the results are plotted.

- Lui [13], Glazer and Hassin [5] and Hassin [8] consider a scheme of auctioning or bribery in an unobservable queue, i.e., at time a customer's need for service arises, he irrevocably either joins the queue or balks. It is not possible for him to observe the queue length before making this decision. In this model, each customer chooses the amount he wishes to pay for priority and then he is placed in the queue ahead of those who paid smaller amounts. It turns out that this decentralized scheme can be used to induce a socially optimal joining rate.

- Myrdal [14] claimed that corrupt officials may deliberately cause administrative delays in service so as to attract more bribes. Lui [13] referred to this claim as Myrdal's hypothesis, and argued that the hypothesis is not always true. For example, if increasing the rate of service is costly to the server, then without a bribe the server has no incentive to supply service, and bribes induce faster service. However, Hassin [8] compared the service rate chosen by a profit maximizer to the socially optimal rate, showing that from this point of view Myrdal's hypothesis is correct. In this paper we show that when the service is slower, i.e., μ is lower, then the server's profit is higher.

- Rosenblum [16] explores a market model where customers trade queue positions. The result is that the customers will be served in decreasing order of value of time, which is known the socially optimal order. But there is a strong assumption in this model that customers do not consider profits that might be gained from transactions in the future, but consider only the reward they receive for the service and their cost of waiting. This model is a kind of overtaking model, but as opposed to our model a customer overtakes other customers only if both he and the overtaken customers agree to this overtaking.

- Hassin, Puerto and Fernandez [10] consider a relative priority approach, where the priority given to a class also depends on state variables associated with other classes. They show that relative priority in an n-class queueing system can reduce server and customer costs. This property is demonstrated in a single server Markovian model where the goal is to minimize a non-linear cost function of class expected waiting times. The priority regimes which they consider are: FCFS, absolute preemptive priorities and DPS (discriminatory processor sharing). Special attention is given to minimizing server's costs when the expected waiting time of each class is restricted.

4 Pure equilibrium strategies

In our model, customers observe the number of customers who are already in the system, and then decide how many customers to overtake. We refer to the number of customers that an arriving customer observes as the total number in the system including the customer in service, but not including the new customer himself. In this section we analyze pure strategies defined by a vector $(k_1, k_2, \ldots)$, where k_i is the number of customers that an arriving customer, who observes i customers in the system, overtakes. Clearly, $k_i \leqslant i$.

Define $f_{i,j}$ to be the expected waiting time of a customer given that there are i customers in front of him (including a customer in service), and j customers behind him and total number of customers in the system is $i+1+j$. In addition define $f_{-1,j} = 0$.

The expected time till either a service completion or a new arrival occurs is $\frac{1}{\lambda+\mu}$. With probability $\frac{\mu}{\lambda+\mu}$ the service completion occurs before a new customer arrives, then his expected waiting time is $f_{i-1,j}$. With probability $\frac{\lambda}{\lambda+\mu}$ a new customer arrives before a service completion occurs, then the new arrival overtakes the present customer if $k_{i+j+1} > j$, and doesn't overtake if $k_{i+j+1} \leqslant j$. Hence, we get

$$f_{i,j} = \frac{1}{\lambda+\mu} + \frac{\mu}{\lambda+\mu} f_{i-1,j} + \frac{\lambda}{\lambda+\mu} f_{i+1,j}, \quad k_{i+j+1} > j, \tag{1a}$$

$$f_{i,j} = \frac{1}{\lambda+\mu} + \frac{\mu}{\lambda+\mu} f_{i-1,j} + \frac{\lambda}{\lambda+\mu} f_{i,j+1}, \quad k_{i+j+1} \leqslant j. \tag{1b}$$

If $k_i \leqslant K$ for all i for some K, this provides boundary conditions, namely $f_{i,j} = \frac{i+1}{\mu}, \forall j \geqslant K$.

If a new customer observes i customers, and decides to overtake k customers, his expected waiting cost is $C_w f_{i-k,k} + kC_o$.

The pure strategy $(k_1, k_2, k_3, \ldots)$ defines an equilibrium if overtaking k_i customers is a best response of a new customer who observes i customers for $i = 1, 2, \ldots$. Therefore, the conditions for equilibrium are:

$C_w f_{i-k_i,k_i} + k_i C_o \leqslant C_w f_{i-k,k} + kC_o$, if $i = 1, 2, \ldots$ and $k = 0, 1, \ldots, i$.

We could not give analytic characterization to the equilibrium strategies. However, we applied numerical analysis to see which strategies are equilibrium for some values of λ, μ and $\frac{C_o}{C_w}$.

We compute strategies $(k_1, k_2, k_3, k_4, k_5, k_6)$ with $k_i = 0, \forall i \geqslant 7$, i.e $7! = 5040$ options. Table 1 contains a list of all strategies such that at least for some values of the input parameters λ, μ, and $\frac{C_o}{C_w}$ they define an equilibrium. In particular, our study shows that even strategies like - $(1, 0, 0, 4, 3, 0)$ or $(0, 2, 0, 0, 5, 5)$ can be equilibrium strategies.

For example, Figure 1 shows the values of $\left(\lambda, \frac{C_w}{C_o}\right)$ for which the strategies $(0, 2, 0, 0, 5, 5)$, $(1, 0, 0, 4, 3, 0)$, $(1, 0, 3, 3, 0, 0)$, and $(1, 2, 3, 4, 4, 0)$ are equilibrium.

There is a detailed example in Appendix-A.

k_1	k_2	k_3	k_4	k_5	k_6
0	0	0	0	0	0,6
0	0	0	0	1	1
0	0	0	0	2	2
0	0	0	0	3	3
0	0	0	0	4	3,4
0	0	0	0	5	0,4,5,6
0	0	0	1	1	0,1
0	0	0	2	2	0,1,2
0	0	0	3	3	0,2,3
0	0	0	4	0	0,6
0	0	0	4	3	0
0	0	0	4	4	0,3,4
0	0	0	4	5	0,5
0	0	1	1	0	0
0	0	1	1	1	0,1
0	0	2	2	0 ,1	0
0	0	2	2	2	0,1,2
0	0	2	2	3	3
0	0	3	3	3	1,2,3
0	0	3	3	4	4
0	0	3	4,0	0	0
0	0	3	0	5	5
0	0	3	3	0,2,3	0
0	0	3	4	4	0,3,4
0	0	3	4	5	5
0	1	1	0,1	0	0
0	1	1	1	1	0,1
0	1	1	2	2	2
0	2	0	0	0	0
0	2	0	0	5	5,6
0	2	0	4	4	0,4
0	2	0	4	5	5
0	2	2	0 ,1	0	0
0	2	2	2	0,1	0
0	2	2	2	2	0,1,2
0	2	2	3	3	0,3
0	2	2	4	4	4
0	2	3	3	0	0
0	2	3	3	3	0,1,2,3
0	2	3	4	4	0,4
0	2	3	4	5	5
1	0	0	0	0	0
1	0	0	0	4	3,4
1	0	0	0	5	4,5,6
1	0	0	4	3	0
1	0	0	4	4	0,3,4
1	0	0	4	5	0,5
1	0	3	3	0	0
1	0	3	3	3	0,2,3
1	0	3	4	4	4
1	0	3	4	5	5
1	1	0	0	0	0
1	1	0	0	2	2
1	1	1	0,1	0	0
1	1	1	1	1	0,1
1	1	2	2	2	0,2
1	1	3	3	3	0,2,3
1	1	3	4	4	4
1	2	0	0	0	0
1	2	2	0	0	0
1	2	2	2	0,1	0
1	2	2	2	2	0,1,2
1	2	2	3	3	3
1	2	2	4	4	4
1	2	3	3	0	0
1	2	3	3	3	0,2,3
1	2	3	4	4	0,4

Table 1: These strategies are equilibrium strategies for some values of λ, μ and $\frac{C_o}{C_w}$.

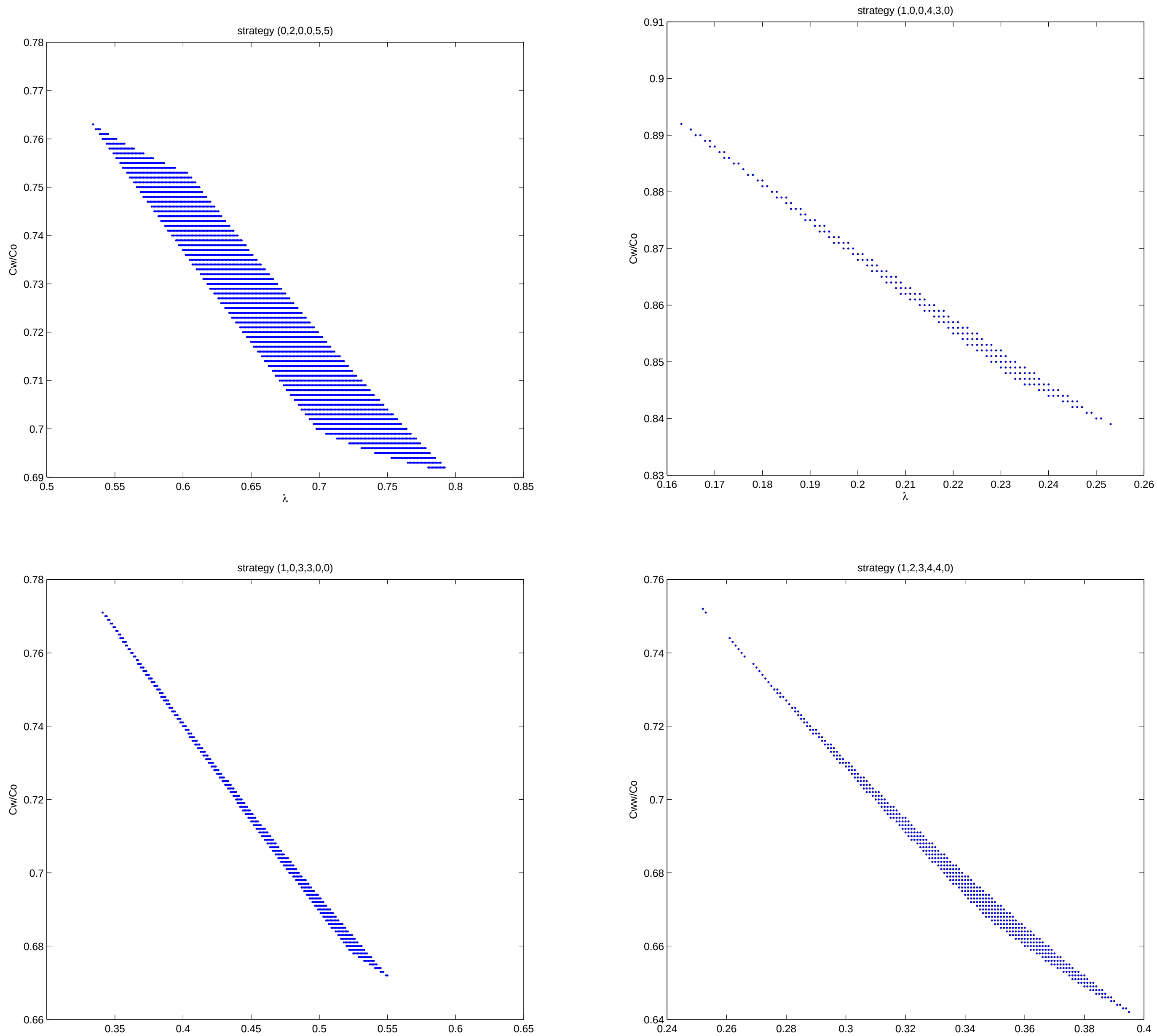

Figure 1: The graphs represent the region of $\left(\lambda, \frac{C_w}{C_o}\right)$ in which the prescribed strategy defines an equilibrium. $\mu = 1$, $C_o = 1$.

5 Overtaking k customers

In this section we consider strategies of the form $k_i = \min\{k, i\}$, i.e., overtaking k customers if there are at least k customers, and overtaking all of them otherwise. Denote this strategy by Σ_k. We observe that if the strategy of all customers is Σ_0, i.e., not overtaking others, then from the assumption of our model that $\frac{C_w}{\mu} < C_o$, the best response of a new customer is also not overtaking. Therefore Σ_0 is always an equilibrium. In addition, we show that Σ_∞, or equivalently, $k_i = i$ for all i is a unique equilibrium when $C_o < \frac{C_w}{\mu}$, Σ_i, $i = 0, 1, \ldots$ is an equilibrium when $\frac{C_w}{\mu} \leqslant C_o \leqslant \frac{C_w}{\mu-\lambda}$, otherwise Σ_0 is a unique equilibrium, see Figure 2.

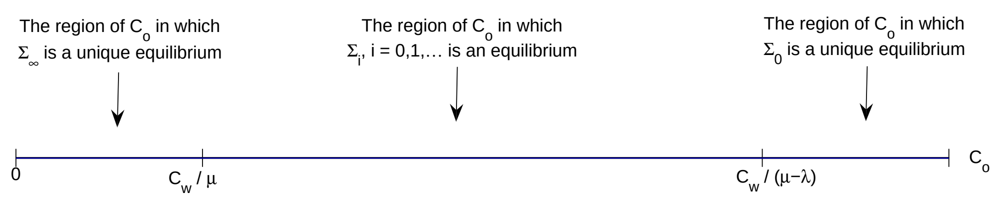

Figure 2: The region of C_o in which the Σ_i, $i = 0, 1, \ldots$ is an equilibrium or a unique equilibrium.

Theorem 5.1 *The strategy Σ_k, $k = 1, 2, \ldots$ defines an equilibrium if and only if*

$$\frac{C_o}{C_w} \leqslant \frac{1}{\mu - \lambda}. \tag{2}$$

Proof: We divide the proof into two parts.

- Suppose that a new customer observes $j \geqslant k$ customers. By overtaking k, he guarantees his place in the queue, because behind him there are k customers, and only they will be overtaken by new customers. Overtaking any additional customer costs C_o and saves $\frac{C_w}{\mu}$. By assumption $C_o > \frac{C_w}{\mu}$ there is no reason to overtake more than k customers.

If he overtakes k customers, his expected cost is

$$C_w \frac{j+1-k}{\mu} + kC_o. \tag{3}$$

Otherwise, if he overtakes i customers, $i < k$, all future customers overtake him till he finishes his service and leaves the system. Therefore his expected waiting time is $j+1-i$ busy periods, and his expected cost is

$$C_w \frac{j+1-i}{\mu-\lambda} + iC_o. \tag{4}$$

The strategy defines an equilibrium if and only if overtaking k customers is a best response of a new customer. Hence, $C_w \frac{j+1-k}{\mu} + kC_o \leqslant C_w \frac{j+1-i}{\mu-\lambda} + iC_o$, or $\frac{C_o}{C_w} \leqslant \frac{1}{\mu-\lambda} + \frac{\lambda(j+1-k)}{\mu(\mu-\lambda)(k-i)}$ for $i = 0, 1, \ldots, k-1$ and $j = k, k+1, \ldots$. The minimum of $\left\{\frac{1}{\mu-\lambda} + \frac{\lambda(j+1-k)}{\mu(\mu-\lambda)(k-i)}\right\}$ over $i = 0, 1, \ldots, k-1$ and $j = k, k+1, \ldots$ is obtained at $i = 0$ and $j = k$. Therefore the condition is $\frac{C_o}{C_w} \leqslant \frac{1}{\mu-\lambda} + \frac{\lambda}{\mu(\mu-\lambda)k}$.

- Suppose that a new customer observes $j = 1, 2, \ldots, k-1$ customers, and chooses overtaking all of them. His expected cost is $C_w \frac{1}{\mu-\lambda} + jC_o$. Otherwise, if he chooses overtaking i customers, $i = 0, 1, \ldots, j-1$, his expected waiting time is $j+1-i$ busy periods, and his expected cost is $C_w \frac{j+1-i}{\mu-\lambda} + iC_o$. In equilibrium overtaking all customers in the queue should be a best response of a new customer. Therefore $C_w \frac{1}{\mu-\lambda} + jC_o \leqslant C_w \frac{j+1-i}{\mu-\lambda} + iC_o$ for $j = 1, 2, \ldots, k-1$ and $i = 0, 1, \ldots, j-1$, or $\frac{C_o}{C_w} \leqslant \frac{1}{\mu-\lambda}$.

■

5.1 Overtaking k customers -Is it worthwhile to observe a longer queue?

One of the interesting questions about Σ_k strategy is: Is it worthwhile to observe a longer queue?

The Σ_k strategy enables a customer who observes at least k customers upon arrival to overtake k of them, and by that to ensure that all future customers will not overtake him. In contrast, a customer who observes less than k customers upon arrival cannot ensure that. He just can overtake all present customers, but all future customers will overtake him till his service completion. We find that there are input parameters for which a customer prefers to observe a longer queue.

Denote the number of observed customers by j.

Theorem 5.2 *Suppose that* $\frac{1}{\mu} \leqslant \frac{C_o}{C_w} \leqslant \frac{1}{\mu-\lambda}$ *and that all customers follow the* Σ_k *strategy. Then:*

1. *The expected cost as a function of* j *is built from two linear functions, one for* $j < k$*, and the second for* $j \geqslant k$*.*

2. *If* $\frac{\lambda}{\mu(\mu-\lambda)} \leqslant \frac{C_o}{C_w}$*, then the the function is monotone increasing for any* j *(Figure 3-**a**).*

3. *If* $\frac{C_o}{C_w} < \frac{\lambda}{k\mu(\mu-\lambda)}$*, then the global minimum is at* k *(Figure 3-**b**). Otherwise, if* $\frac{\lambda}{k\mu(\mu-\lambda)} < \frac{C_o}{C_w}$*, then the global minimum is at* 0*.*

4. *If* $\frac{C_o}{C_w} < \frac{\lambda-(\mu-\lambda)(j-k)}{\mu(\mu-\lambda)(k-j')}$*, when* $j \geqslant k$ *and* $j' < k$*, then a new customer prefers to observe a longer queue, i.e.,* j *grater than smaller (Figure 3-**c**).*

Proof:

- Suppose that a new customer observes $j \geqslant k$ customers. Then according to Σ_k he overtakes k of them. Future arrivals customers do not overtake him, and therefore, his expected cost is $C_w\frac{j+1-k}{\mu} + kC_o$.

- Suppose that a new customer observes $j < k$ customers. Then according to Σ_k he chooses overtaking all of them. Future customers overtake him, and therefore, his expected cost is $C_w\frac{1}{\mu-\lambda} + jC_o$.

Hence, the expected cost of a new customer is

$$C_w\frac{j+1-k}{\mu} + kC_o, \quad j \geqslant k$$

,

$$C_w \frac{1}{\mu - \lambda} + jC_o, \quad j < k$$

.

The functions $C_w \frac{j+1-k}{\mu} + kC_o$ and $C_w \frac{1}{\mu-\lambda} + jC_o$ are both monotone increasing in j. If $C_w \frac{1}{\mu-\lambda} + j'C_o \leqslant C_w \frac{j+1-k}{\mu} + kC_o$ when $j' = k-1$ and $j = k$, or equivalently $\frac{\lambda}{\mu(\mu-\lambda)} \leqslant \frac{C_o}{C_w}$ then the expected cost of a new customer as a function of the number of customers in the system is a monotone and increasing. Otherwise, this function is built from two monotone increasing functions with a break-point at k.

Since both functions are monotone increasing then k is a global minimum if a new customer prefers to observe k rather than an empty queue, i.e., $C_w \frac{1}{\mu} + kC_o < C_w \frac{1}{\mu-\lambda}$, or equivalently, $\frac{C_o}{C_w} < \frac{\lambda}{k\mu(\mu-\lambda)}$. If a new customer prefers to observe an empty queue rather than k then 0 is a global minimum and then $C_w \frac{1}{\mu-\lambda} < C_w \frac{1}{\mu} + kC_o$, or equivalently, $\frac{\lambda}{k\mu(\mu-\lambda)} < \frac{C_o}{C_w}$.

A new customer prefers to observe a longer queue, i.e., prefers to observe $j \geqslant k$ customers rather than $j' < k$, if $C_w \frac{j+1-k}{\mu} + kC_o < C_w \frac{1}{\mu-\lambda} + j'C_o$, or equivalently if $\frac{C_o}{C_w} < \frac{\lambda-(\mu-\lambda)(j-k)}{\mu(\mu-\lambda)(k-j')}$. ∎

For example, see Figure 3. In figure ***a*** the function is monotone increasing which means that an arriving customer always prefers to observe a shorter queue. In figure ***b*** an arriving customer prefers to observe $k = 4$ rather than 0 customers in the system. If he observes $k = 4$ customers, then he overtakes all of them and all future arrival customers do not overtake him, and his expected cost is $\frac{C_w}{\mu} + kC_o = 5$. Otherwise, if he observes 0 customers, then all future arrival customers overtake him, and his expected cost is $\frac{C_w}{\mu-\lambda} = 8$. In conclusion, in this case we get that $\frac{C_w}{\mu} + kC_o = 5 < \frac{C_w}{\mu-\lambda} = 8$, i.e., it is worthwhile to observe a longer queue. In figure ***c*** an arriving customer prefers to observe 0 rater than $k = 12$ customers in the system. If he observes $k = 12$ customers, then he overtakes all of them and all future arrival customers do not overtake him, and his expected cost is $\frac{C_w}{\mu} + kC_o = 13$.

Otherwise, If he observes 0 customers, then all of them and all future arrival customers overtake him, and his expected cost is $\frac{C_w}{\mu-\lambda} = 8$. In conclusion in this case we get that $\frac{C_w}{\mu} + kC_o = 13 > \frac{C_w}{\mu-\lambda} = 8$, i.e., it is worthwhile to observe a shorter queue.

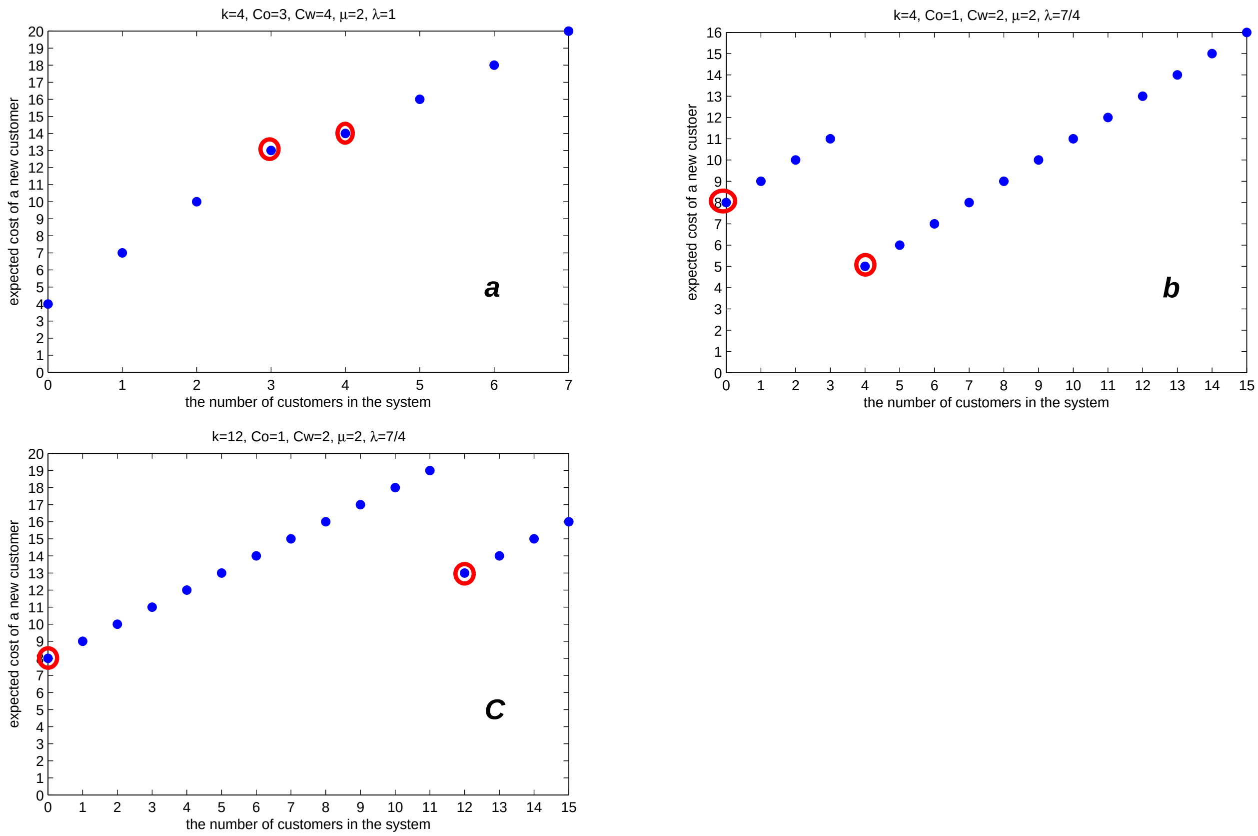

Figure 3: Expected cost as a function of the number of observed customers.

5.2 Overtaking k customers - two actions mixed strategy

In this section we consider the mixed strategy $\Sigma_{k,\mathbf{p}}$. The mixed strategy $\Sigma_{k,\mathbf{p}}$ is defined as follows: For a given integer $k \geqslant 1$ and a vector $\mathbf{p} = (p_k, p_{k+1}, \ldots)$ such that $p_i \in [0,1]$ for every $i = k, \ldots$, a customer who observes upon arrival $i \geq k$ customers in the system (including the one in service) overtakes k customers with probability p_i and $k-1$ customers otherwise. If there are at most $k-1$ customers in the system, the customer overtakes them all.

Theorem 5.3 *The mixed strategy* $\Sigma_{k,\mathbf{p}}$ *defines an equilibrium if and only if* $\frac{1}{\mu} \leqslant \frac{C_o}{C_w} \leqslant \frac{1}{\mu-\lambda}$, *and for some* $x \in \left[\max\left\{0, \frac{(\lambda+\mu)^2}{\mu\lambda}\left[\frac{\mu}{\lambda+\mu} - \frac{C_w}{C_o\mu}\right]\right\}, \min\left\{1, \frac{(\lambda+\mu)^2}{\mu\lambda}\left[1 - \frac{C_w}{C_o\mu}\right]\right\}\right]$

$p_k = x,$

$p_{k+1} = \frac{\lambda+\mu}{\lambda}\left[1 - \frac{C_w}{C_o\mu}\right] - \frac{\mu}{\lambda+\mu}x,$

$p_{k+j} = 1 - \frac{C_w}{C_o\mu}, \quad \forall j \geqslant 2.$

In particular, such x *satisfies the following:*

1. $0 \leqslant x \leqslant \frac{(\lambda+\mu)^2}{\mu\lambda}\left[1 - \frac{C_w}{C_o\mu}\right]$, *if* $\frac{1}{\mu} \leqslant \frac{C_o}{C_w} \leqslant \frac{(\lambda+\mu)^2}{\mu(\mu^2+\mu\lambda+\lambda^2)}$.

2. $0 \leqslant x \leqslant 1$, *if* $\frac{(\lambda+\mu)^2}{\mu(\mu^2+\mu\lambda+\lambda^2)} \leqslant \frac{C_o}{C_w} \leqslant \frac{\lambda+\mu}{\mu^2}$.

3. $\frac{(\lambda+\mu)^2}{\mu\lambda}\left[\frac{\mu}{\lambda+\mu} - \frac{C_w}{C_o\mu}\right] \leqslant x \leqslant 1$, *if* $\frac{\lambda+\mu}{\mu^2} \leqslant \frac{C_o}{C_w} \leqslant \min\left\{\frac{1}{\mu-\lambda}, \frac{(\lambda+\mu)^2}{\mu^3}\right\}$.

For example, suppose that $\frac{C_o}{C_w} = \frac{\lambda+\mu}{\mu^2}$, then $p_k = 0$ and $p_{k+1} = 1$.

Proof: We say that a customer is in state (i, j) if there are exactly i customers in front of him (including the one in service) and exactly j customers behind him. We denote by $f_{i,j}$ the expected (residual) waiting time of a customer in state (i, j) given that all future customers adopt the strategy $\Sigma_{k,\mathbf{p}}$. In addition let $f_{-1,j} = 0$.

- Consider a customer in state (i, j) where $j \leq k - 2$. Clearly, all future customers overtake him. Therefore, his waiting time amounts to a total of $i + 1$ busy periods, and

$$f_{i,j} = \frac{i+1}{\mu - \lambda}, \quad \forall j \leq k - 2. \tag{5}$$

- Consider a customer in state $(j - 1, k - 1)$, $j \geqslant 1$. The expected time till either a service completion occurs or a new arrival is $\frac{1}{\lambda+\mu}$. With probability $\frac{\lambda}{\lambda+\mu}$ a new customer arrives before a service completion occurs. The new arrival overtakes k customers with probability p_{j+k-1}, in this case the customer's expected residual waiting time is $f_{j,k-1}$. Otherwise, the new arrival overtakes only $k - 1$ customers and the customer's position is guaranteed, with expected waiting time $f_{j-1,k} = \frac{j}{\mu}$. With probability $\frac{\mu}{\lambda+\mu}$ the service completion occurs before a new arrival, and then the customer's expected residual waiting time is $f_{j-2,k-1}$. Therefore,

$f_{j-1,k-1} = \frac{1}{\lambda+\mu} + \frac{\mu}{\lambda+\mu} f_{j-2,k-1} + \frac{\lambda p_{j+k-1}}{\lambda+\mu} f_{j,k-1} + \frac{\lambda(1-p_{j+k-1})}{\lambda+\mu}\frac{j}{\mu}, \quad j \geqslant 1$

or equivalently,

$$f_{j,k-1} = \frac{1}{\lambda p_{j+k-1}} \left[(\lambda+\mu) f_{j-1,k-1} - \mu f_{j-2,k-1} - 1 - (1-p_{j+k-1}) \frac{\lambda j}{\mu} \right], \quad j \geqslant 1. \tag{6}$$

In particular,

$$f_{1,k-1} = \frac{1}{\lambda p_k} \left[(\mu+\lambda) f_{0,k-1} - 1 - (1-p_k) \frac{\lambda}{\mu} \right]. \tag{7}$$

- Consider now a customer who observes $k+j-1$ customers upon arrival, where $j \geq 1$. If he overtakes k customers, he guarantees his position in the queue and his expected cost is $C_w \frac{j}{\mu} + kC_o$. Otherwise, if he overtakes only $k-1$ customers, his expected cost is $C_w f_{j,k-1} + (k-1)C_o$. For $\Sigma_{k,\mathbf{p}}$ to define an equilibrium strategy, it must be that the customer is indifferent between the two options, hence

$$f_{j,k-1} = \frac{C_o}{C_w} + \frac{j}{\mu}, \quad \forall j \geq 1. \tag{8}$$

Substituting $f_{1,k-1}$ from (8) in (7) gives

$$p_k = \frac{1}{\lambda} \frac{C_w}{C_o} \left[(\lambda+\mu) f_{0,k-1} - 1 - \frac{\lambda}{\mu} \right]. \tag{9}$$

Substituting $f_{1,k-1}$ from (8) in (6) for $j = 2$ gives

$$p_{k+1} = 1 + \frac{\mu}{\lambda} - \frac{1}{\lambda} \frac{C_w}{C_o} \left[\mu f_{0,k-1} + \frac{\lambda}{\mu} \right]. \tag{10}$$

For $j \geqslant 2$, substituting $f_{j,k-1}$, $f_{j-1,k-1}$, and $f_{j-2,k-1}$ from (8) in (6) gives,

$$p_{k+j} = 1 - \frac{C_w}{C_o \mu}, \quad \forall j \geq 2. \tag{11}$$

We denote p_k by x. Substituting p_k from (9), and p_{k+1} from (10), in $p_k + \frac{\lambda+\mu}{\mu} p_{k+1}$ gives

$$p_{k+1} = \frac{\lambda+\mu}{\lambda}\left[1 - \frac{C_w}{C_o\mu}\right] - \frac{\mu}{\lambda+\mu}x. \tag{12}$$

Since $0 \leqslant p_{k+1} \leqslant 1$, we get that $\frac{(\lambda+\mu)^2}{\mu\lambda}\left[\frac{\mu}{\lambda+\mu} - \frac{C_w}{C_o\mu}\right] \leqslant x \leqslant \frac{(\lambda+\mu)^2}{\mu\lambda}\left[1 - \frac{C_w}{C_o\mu}\right]$. $x = p_k$, hence we must get that $0 \leqslant x \leqslant 1$. Therefore,

$$\max\left\{0, \frac{(\lambda+\mu)^2}{\mu\lambda}\left[\frac{\mu}{\lambda+\mu} - \frac{C_w}{C_o\mu}\right]\right\} \leqslant x \leqslant \min\left\{1, \frac{(\lambda+\mu)^2}{\mu\lambda}\left[1 - \frac{C_w}{C_o\mu}\right]\right\}.$$

We consider these cases:

1. $\frac{1}{\mu} \leqslant \frac{C_o}{C_w} \leqslant \frac{(\lambda+\mu)^2}{\mu(\mu^2+\mu\lambda+\lambda^2)}$. In this case $0 \leqslant x \leqslant \frac{(\lambda+\mu)^2}{\mu\lambda}\left[1 - \frac{C_w}{C_o\mu}\right]$.

2. $\frac{(\lambda+\mu)^2}{\mu(\mu^2+\mu\lambda+\lambda^2)} \leqslant \frac{C_o}{C_w} \leqslant \frac{\lambda+\mu}{\mu^2}$. In this case $0 \leqslant x \leqslant 1$.

3. $\frac{\lambda+\mu}{\mu^2} \leqslant \frac{C_o}{C_w} \leqslant \min\left\{\frac{1}{\mu-\lambda}, \frac{(\lambda+\mu)^2}{\mu^3}\right\}$. In this case $\frac{(\lambda+\mu)^2}{\mu\lambda}\left[\frac{\mu}{\lambda+\mu} - \frac{C_w}{C_o\mu}\right] \leqslant x \leqslant 1$.

We now analyze the other equilibrium conditions and show that they are satisfied if and only if $\frac{C_o}{C_w} \leqslant \frac{1}{\mu-\lambda}$.

- The best response of a new customer who observes $j \leq k-2$ customers is overtaking all of them. Hence, $C_w f_{0,j} + C_o j \leqslant C_w f_{j-l,l} + C_o l$, $l = 0, 1, 2, \ldots, j-1$. Substituting $f_{0,j}$ and $f_{j-l,l}$ from (5), this gives

$$\frac{C_0}{C_w} \leqslant \frac{1}{\mu-\lambda}. \tag{13}$$

- Consider the best response for a new customer who observes $k-1$ customers. If he overtakes all of them, his expected cost is $C_w f_{0,k-1} + C_o(k-1)$. Otherwise, if he overtakes only $l \leq k-2$ customers, his expected waiting time is $f_{j-l,l}$, substituting $f_{j-l,l}$ from (5) we get that his expected cost is $C_w \frac{k-l}{\mu-\lambda} + C_o l$. In equilibrium the best response is overtaking all customers, hence $C_w f_{0,k-1} + C_o(k-1) \leqslant C_w \frac{k-l}{\mu-\lambda} + C_o l$, or $\frac{C_0}{C_w} \leqslant \frac{k-l}{(\mu-\lambda)(k-1-l)} - \frac{f_{0,k-1}}{k-1-l}$. $f_{0,k-1}$ is bounded from above by the expected length of a busy period, because it is the maximum time till a customer in service leaves the system, even if all new arrival customers overtake him. Therefore, $\frac{k-l}{(\mu-\lambda)(k-1-l)} - \frac{f_{0,k-1}}{k-1-l} \geqslant \frac{k-l}{(\mu-\lambda)(k-1-l)} - \frac{1}{\mu-\lambda}\frac{1}{k-1-l} = \frac{1}{\mu-\lambda}$, and we get the condition (13).

- The last case that we should check is if a new customer observes $j \geqslant k$ customers and chooses overtaking only $m \leq k-2$ customers, then his expected cost is $C_w f_{j-m,m} + mC_o$. In a symmetric equilibrium the best response is overtaking k customers and not less. Therefore $C_w f_{j-m,m} + mC_o \geqslant C_w \frac{j-k+1}{\mu} + kC_o$, or $f_{j-m,m} \geqslant \frac{C_o}{C_w}(k-m) + \frac{j-k+1}{\mu}$. Substituting $f_{j-m,m}$ from (5) we get, $\frac{j-m+1}{\mu-\lambda} \geqslant \frac{C_o}{C_w}(k-m) + \frac{j-k+1}{\mu}$, or
$\frac{C_o}{C_w} \leqslant \frac{1}{\mu-\lambda} + \frac{\lambda(j-k+1)}{\mu(\mu-\lambda)}\frac{1}{k-m}$, and $\frac{1}{\mu-\lambda} + \frac{\lambda(j-k+1)}{\mu(\mu-\lambda)}\frac{1}{k-m} > \frac{1}{\mu-\lambda}$. Therefore we get the condition (13).

■

Observation 5.4 Theorem 5.3 shows that for any $k = 1, 2, \ldots\ p_{k+j} = 1 - \frac{C_w}{C_o\mu}, \quad \forall j \geq 2$. There is a different approach for calculating p_j when $j \to \infty$. A customer who observes $j+k-1$ customers, and overtakes $k-1$ of them will be overtaken till a new arrival chooses to not overtake k customers with probability $1 - p_{j+k}$. In other words, the time till the choice of an arriving customer is not overtaking has a geometric distribution, with probability $1 - p_{j+k}$ for success, and probability p_{j+k} for failure. Hence, the number of customers who arrive till the choice of the arriving customer is not overtaking, is $\frac{1}{1-p_{j+k}} - 1$ (not including the customer who chooses not to overtake). Therefore, the residual expected waiting time of a customer who observes $j+k-1$ customers, and overtakes $k-1$ of them, consists of the service times of all customers who arrive till the first time he isn't overtaken, plus j service times of customers that were before him, plus one service time of himself. Hence, when $j \to \infty$ $f_{j,k-1} = \frac{j + \left(\frac{1}{1-p_{j+k}} - 1\right) + 1}{\mu} = \frac{j + \frac{1}{1-p_{j+k}}}{\mu}$. Substituting $f_{j,k-1}$ from (8) gives when $j \to \infty$ $p_j = 1 - \frac{C_w}{C_o\mu}$ which is proved in Theorem 5.3 .

A (symmetric) equilibrium strategy is, by definition, a best response against itself. However, it need not to be the unique best response. Specifically, let y be an equilibrium strategy. There may be a best response strategy $z \neq y$ such that z is strictly a better response against itself than y is. In this case, y is unstable in the sense that when starting with y, it may be that the players adopt the best response z, and then a new equilibrium, at z, will be reached. If no such z exists then y is said to be an evolutionarily stable strategy or ESS. Note that if y is an equilibrium strategy and it is the unique best response against itself, then it is necessarily ESS.

Theorem 5.5 *The equilibrium mixed strategy $\Sigma_{k,\mathbf{p}}$ is not ESS.*

Proof: Denote the equilibrium mixed strategy $\Sigma_{k,\mathbf{p}}$ by y. Consider a strategy $z = \sum_{k-1} = (1, 2, \ldots, k-1, k-1, \ldots)$. In Section 5 we show that this is an equilibrium strategy, and now we show that this strategy is strictly a better response against itself than y is, i.e., if all customers use z then z is the best response. If an arriving customer observes $j \geqslant k-1$ customers, then from the assumption of our model that $\frac{C_w}{\mu} < C_o$ his best response is to overtake only $k-1$ customers because then all future customers do not over take him. Therefore, z is strictly a better response against itself than y is. ∎

5.3 Overtaking k customers - three actions mixed strategy

In this section we check whether there is an equilibrium strategy, where customers are indifferent between overtaking k, $k-1$ or $k-2$ customers.

Now the mixed strategy $\Sigma_{k,\mathbf{p}}$ is defined as follows: For a given integer $k \geqslant 1$ and a matrix $\mathbf{p} = \begin{pmatrix} p_{k-1}^{k-1} & 0 \\ p_k^{k-1} & p_k^k \\ p_{k+1}^{k-1} & p_{k+1}^k \\ \cdot & \cdot \\ \cdot & \cdot \end{pmatrix}$ such that $p_i^{k-1}, p_j^k \in [0,1]$ for every $i = k-1, \ldots$ and $j = k, \ldots$. A customer who observes upon arrival $i \geq k$ customers in the system (including the one in service) overtakes k customers with probability p_i^k, $k-1$ customers with probability p_i^{k-1}, and $k-2$ customers with probability $p_i^{k-2} = 1 - p_i^k - p_i^{k-1}$. A customer who observes upon arrival $k-1$ customers in the system (including the one in service) overtakes $k-1$ customers with probability p_{k-1}^{k-1}, and $k-2$ customers otherwise. If there are at most $k-2$ customers in the system, the customer overtakes them all.

Theorem 5.6 *The mixed strategy $\Sigma_{k,\mathbf{p}}$ defines a unique equilibrium where customers are indifferent between overtaking k, $k-1$ or $k-2$ customers if and only if $\frac{1}{\mu} \leqslant \frac{C_o}{C_w} \leqslant \frac{1}{\mu-\lambda}$,*

$p_k^k = x,\ p_{k-1}^{k-1} = y,\ p_k^{k-1} = z,$

$p_{k+1}^k = \frac{\lambda+\mu}{\lambda}\left[1 - \frac{C_w}{C_o\mu}\right] - \frac{\mu}{\lambda+\mu}x,$

$p_{k+1}^{k-1} = \frac{\mu^2(\lambda+\mu)^2 - \frac{C_w}{C_o}\mu(\lambda+\mu)^2 - \mu^2\lambda(\lambda x + (\lambda+\mu)y)}{\lambda(\lambda+\mu)(\lambda^2+\mu\lambda+\mu^2)} - \frac{\mu(\lambda+\mu)}{\lambda^2+\mu\lambda+\mu^2}z,$

$p_{k+j}^k = 1 - \frac{C_w}{C_o\mu}, \quad \forall j \geqslant 2,$

$p_{k+j}^{k-1} = 0 \quad \forall j \geqslant 2,$

$0 \leqslant x \leqslant 1,\ 0 \leqslant y \leqslant 1,\ 0 \leqslant z \leqslant 1,\ 0 \leqslant p_{k+1}^k \leqslant 1,\ 0 \leqslant p_{k+1}^{k-1} \leqslant 1,$

$x + z \leqslant 1,$

$p_{k+j}^{k-1} + p_{k+j}^k \leqslant 1 \quad \forall j \geqslant 1.$

Proof: We say that a customer is in state (i, j) if there are exactly i customers in front of him (including the one in service) and exactly j customers behind him. Define $f_{i,j}$ to be the expected waiting time of a customer in state (i, j) when all future customers adopt the strategy $\Sigma_{k,\mathbf{p}}$. In addition let $f_{-1,j} = 0$.

- Consider a customer who observes $k+j-1$ customers upon arrival, where $j \geq 1$. If he overtakes k customers, he guarantees his position in the queue and his expected cost is $C_w\frac{j}{\mu} + kC_o$. Otherwise, if he overtakes $k-1$ customers, his expected cost is $C_w f_{j,k-1} + (k-1)C_o$. Finally, if he overtakes $k-2$ customers, his expected cost is $C_w f_{j+1,k-2} + (k-2)C_o$. By assumption the customer is indifferent between the three options, hence

$$f_{j,k-1} = \frac{C_o}{C_w} + \frac{j}{\mu}, \quad \forall j \geq 1, \tag{14}$$

and,

$$f_{j,k-2} = 2\frac{C_o}{C_w} + \frac{j-1}{\mu}, \quad \forall j \geq 2. \tag{15}$$

- Consider a customer who observes $k-1$ customers upon arrival. If he overtakes $k-1$ customers, his expected cost is $C_w f_{0,k-1} + (k-1)C_o$. Otherwise, if he overtakes $k-2$ customers, his expected cost is $C_w f_{1,k-2} + (k-2)C_o$. By assumption the customer is indifferent between the two options, hence

$$f_{0,k-1} = f_{1,k-2} - \frac{C_o}{C_w}. \tag{16}$$

- Consider a customer who observes $k-j$, $j \geqslant 2$ customers upon arrival. If he overtakes all of them then his waiting time is equivalent to a busy period and his expected cost is $C_w f_{0,k-j} + (k-j)C_o$, or equivalently $C_w \frac{1}{\mu-\lambda} + (k-j)C_o$. Otherwise, if he overtakes only $i < k-j$ customers, his waiting time is equivalent to $k-j-i+1$ busy periods and his expected cost is $C_w f_{k-j-i,i} + iC_o$, or equivalently $C_w \frac{k-j-i+1}{\mu-\lambda} + iC_o$. The strategy defines an equilibrium if and only if overtaking all customers is a best response of a new customer. Hence, $C_w \frac{1}{\mu-\lambda} + (k-j)C_o < C_w \frac{k-j-i+1}{\mu-\lambda} + iC_o$, or equivalently $\frac{C_o}{C_w} \leqslant \frac{1}{\mu-\lambda}$.

- Consider a customer in state $(j-1, k-1)$, $j \geqslant 1$. The expected time till the next arrival or service completion occurs is $\frac{1}{\lambda+\mu}$. With probability $\frac{\lambda}{\lambda+\mu}$ a new customer arrives before a service completion occurs. The new arrival overtakes k customers with probability p^k_{j+k-1}, in this case the customer's expected residual waiting time is $f_{j,k-1}$. Otherwise, if the new arrival overtakes only $k-1$ customers (with probability p^{k-1}_{j+k-1}), or $k-2$ customers (with probability $1-p^k_{j+k-1}-p^{k-1}_{j+k-1}$), the customer's position is guaranteed, with expected waiting time $f_{j-1,k} = \frac{j}{\mu}$. With probability $\frac{\mu}{\lambda+\mu}$ the service completion occurs before a new arrival, and then the customer's expected residual waiting time is $f_{j-2,k-1}$. Therefore, for $j \geqslant 1$

$f_{j-1,k-1} = \frac{1}{\lambda+\mu} + \frac{\mu}{\lambda+\mu} f_{j-2,k-1} + \frac{\lambda p^k_{j+k-1}}{\lambda+\mu} f_{j,k-1} + \frac{\lambda(1-p^k_{j+k-1})}{\lambda+\mu} \frac{j}{\mu}$,

or equivalently,

$$f_{j,k-1} = \frac{1}{\lambda p^k_{j+k-1}} \left[(\lambda+\mu) f_{j-1,k-1} - \mu f_{j-2,k-1} - 1 - \frac{\lambda j}{\mu} \left(1 - p^k_{j+k-1} \right) \right], \quad j \geqslant 1. \tag{17}$$

Substituting $f_{1,k-1}$ from (14) in (17) for $j = 1$ gives

$$p^k_k = \frac{1}{\lambda} \frac{C_w}{C_o} \left[(\lambda+\mu) f_{0,k-1} - 1 - \frac{\lambda}{\mu} \right]. \tag{18}$$

Substituting $f_{1,k-1}$ and $f_{2,k-1}$ from (14) in (17) for $j = 2$ gives

$$p^k_{k+1} = 1 + \frac{\mu}{\lambda} - \frac{1}{\lambda} \frac{C_w}{C_o} \left[\mu f_{0,k-1} + \frac{\lambda}{\mu} \right]. \tag{19}$$

For $j > 2$, substituting $f_{j,k-1}$, $f_{j-1,k-1}$, and $f_{j-2,k-1}$ from (14) in (17) gives,

$$p^k_{k+j} = 1 - \frac{C_w}{C_o \mu}, \quad \forall j \geq 2 \tag{20}$$

We denote p^k_k by x. Substituting $f_{0,k-1}$ from (18) in (19) gives

$$p^k_{k+1} = \frac{\lambda+\mu}{\lambda} \left[1 - \frac{C_w}{C_o \mu} \right] - \frac{\mu}{\lambda+\mu} x. \tag{21}$$

- Consider a customer in state $(j-1, k-2)$, $j \geqslant 2$. The expected time till the next arrival or service completion occurs is $\frac{1}{\lambda+\mu}$. With probability $\frac{\lambda}{\lambda+\mu}$ a new customer arrives before a service completion occurs. The new arrival overtakes k customers with probability p^k_{j+k-2} and $k-1$ customers with probability p^{k-1}_{j+k-2}, in both cases the customer's expected residual waiting time is $f_{j,k-2}$. The new arrival overtakes $k-2$ customers with probability $1 - p^k_{j+k-2} - p^{k-1}_{j+k-2}$, in this case customer's expected residual waiting time is $f_{j-1,k-1}$. With probability $\frac{\mu}{\lambda+\mu}$ the service completion occurs before a new arrival, and then the customer's expected residual waiting time is $f_{j-2,k-2}$. Therefore, for $j \geqslant 2$

$f_{j-1,k-2} = \frac{1}{\lambda+\mu} + \frac{\mu}{\lambda+\mu} f_{j-2,k-2} + \frac{\lambda(p^k_{j+k-2}+p^{k-1}_{j+k-2})}{\lambda+\mu} f_{j,k-2} + \frac{\lambda(1-p^k_{j+k-2}-p^{k-1}_{j+k-2})}{\lambda+\mu} f_{j-1,k-1},$

or equivalently,

$$f_{j,k-2} = \frac{1}{\lambda(p^k_{j+k-2}+p^{k-1}_{j+k-2})} \times \left[(\lambda+\mu) f_{j-1,k-2} - \mu f_{j-2,k-2} - 1 - \lambda\left(1 - p^k_{j+k-2} - p^{k-1}_{j+k-2}\right) f_{j-1,k-1}\right], \quad j \geqslant 2. \tag{22}$$

- Consider a customer in state $(1, k-2)$. The expected time till the next arrival or service completion occurs is $\frac{1}{\lambda+\mu}$. With probability $\frac{\lambda}{\lambda+\mu}$ a new customer arrives before a service completion occurs. The new arrival overtakes k customers with probability p^k_k and overtakes $k-1$ customers with probability p^{k-1}_k, in both cases the customer's expected residual waiting time is $f_{2,k-2}$. The new arrival overtakes $k-2$ customers with probability $1 - p^k_k - p^{k-1}_k$, in this case customer's expected residual waiting time is $f_{1,k-1}$. With probability $\frac{\mu}{\lambda+\mu}$ the service completion occurs before a new arrival, and then the customer's expected residual waiting time is $f_{0,k-2}$. Therefore,

  $f_{1,k-2} = \frac{1}{\lambda+\mu} + \frac{\mu}{\lambda+\mu} f_{0,k-2} + \frac{\lambda(p^k_k+p^{k-1}_k)}{\lambda+\mu} f_{2,k-2} + \frac{\lambda(1-p^k_k-p^{k-1}_k)}{\lambda+\mu} f_{1,k-1}$. Substituting $f_{2,k-2}$ from (15) and $f_{1,k-1}$ from (14) and $p^k_k = x$ gives

$$f_{1,k-2} = \frac{1}{\mu} + \frac{\mu}{\lambda+\mu} f_{0,k-2} + \frac{\lambda}{\lambda+\mu} \frac{C_o}{C_w} \left(x + 1 + p^{k-1}_k\right). \tag{23}$$

- Consider a customer in state $(0, k-2)$. The expected time till the next arrival or service completion occurs is $\frac{1}{\lambda+\mu}$. With probability $\frac{\lambda}{\lambda+\mu}$ a new customer arrives before a service completion occurs. The new arrival overtakes $k-1$ customers with probability p^{k-1}_{k-1}, in this case customer's expected residual waiting time is $f_{1,k-2}$, and overtakes $k-2$ customers with probability $1 - p^{k-1}_{k-1}$, in this case customer's expected residual waiting time is $f_{0,k-1}$. Therefore, $f_{0,k-2} = \frac{1}{\lambda+\mu} + \frac{\lambda p^{k-1}_{k-1}}{\lambda+\mu} f_{1,k-2} + \frac{\lambda(1-p^{k-1}_{k-1})}{\lambda+\mu} f_{0,k-1}$.

Substituting $f_{0,k-1}$ from (16) gives

$$f_{0,k-2} = \frac{1}{\lambda+\mu} + \frac{\lambda}{\lambda+\mu} f_{1,k-2} - \frac{\lambda}{\lambda+\mu}\frac{C_o}{C_w}\left(1 - p_{k-1}^{k-1}\right). \tag{24}$$

Substituting $f_{2,k-2}$ from (15), $f_{1,k-1}$ from (14), $p_k^k = x$, and $j = 2$ in (22) gives

$$p_k^{k-1} = \frac{C_w}{C_o\lambda}\left[(\lambda+\mu) f_{1,k-2} - \mu f_{0,k-2} - 1 - \frac{\lambda}{\mu}\right] - 1 - x. \tag{25}$$

Substituting $f_{3,k-2}$ and $f_{2,k-2}$ from (15), $f_{2,k-1}$ from (14), p_{k+1}^k from (21), and $j = 3$ in (22) gives

$$p_{k+1}^{k-1} = \frac{\mu}{\lambda} + \frac{\mu}{\lambda+\mu} x - \frac{C_w}{C_o\lambda}\left[\mu f_{1,k-2} - 1\right]. \tag{26}$$

Substituting $f_{j,k-2}$,$f_{j-1,k-2}$,$f_{j-2,k-2}$ from (15) and $f_{j-1,k-1}$ from (14) in (22) gives $p_{k+j}^{k-1} = 1 - p_{k+j}^k - \frac{C_w}{C_o\mu}, \forall j \geq 2$, and substituting p_{k+j}^k from (20) gives

$$p_{k+j}^{k-1} = 0, \quad \forall j \geq 2. \tag{27}$$

We denote p_{k-1}^{k-1} by y and p_k^{k-1} by z.

Substituting $f_{1,k-2}$ from (23) in (24) gives

$$f_{0,k-2} = \frac{1}{\lambda^2 + \lambda\mu + \mu^2}\left[\frac{(\lambda+\mu)^2}{\mu} + \frac{C_o}{C_w}\left(\lambda^2(x+1+z) - \lambda(\lambda+\mu)(1-y)\right)\right]. \tag{28}$$

Substituting this $f_{0,k-2}$ from (28) in (25) gives

$$\begin{aligned} z = {} & \tfrac{C_w}{C_o}\left[(\lambda+\mu) f_{1,k-2} - \tfrac{(\lambda+\mu)^2}{\lambda^2+\lambda\mu+\mu^2} - 1 - \tfrac{\lambda}{\mu}\right] - \\ & 1 - x - \tfrac{\mu}{\lambda(\lambda^2+\lambda\mu+\mu^2)}\left[\lambda^2(x+1+z) - \lambda(\lambda+\mu)(1-y)\right]. \end{aligned} \tag{29}$$

Substituting $f_{1,k-2}$ from (29) in (26) gives

$p_{k+1}^{k-1} = \frac{\mu^2(\lambda^2+\mu\lambda+\mu^2) - \frac{C_w}{C_o}\mu(\lambda+\mu)^2 - \mu^2\lambda(\lambda x + (\lambda+\mu)y - \mu)}{\lambda(\lambda+\mu)(\lambda^2+\mu\lambda+\lambda^2)} - \frac{\mu(\lambda+\mu)}{\lambda^2+\mu\lambda+\mu^2} z.$

$0 \leqslant x \leqslant 1$, $0 \leqslant y \leqslant 1$, $0 \leqslant z \leqslant 1$, $0 \leqslant p_{k+1}^k \leqslant 1$, $0 \leqslant p_{k+1}^{k-1} \leqslant 1$, because they are all probabilities.

In addition $p_i^{k-1} + p_i^k \leqslant 1$ for every $i = k, \ldots,$ because if a customer observes i customers, then the sum of probabilities for overtaking k or $k-1$ not greater than one (the sum is equal to one for the sum of all three probabilities, for overtaking k, $k-1$ and $k-2$ customers).

■

The equilibrium conditions are not an empty range. A detailed example is given in Appendix-B.

6 Profit maximization

In this section we compare two models. In both of them the customers purchase priority, customers are identical except their arrival time, there is no balking or reneging, and decisions are made upon arrival and cannot be changed later. The service disciplines are preemptive resume. We compare the server's maximum expected profit *per customer* under equilibrium conditions.

6.1 Maximum profit in the current priority (CP) model

The first model is the model of Section 2. In this model purchasing priority enables overtaking present customers. Upon arrival, a new customer decides on the number of current customers that he overtakes, and pays per each overtaken customer a fixed cost. In this model an arriving customer overtakes customers who are currently in the system, but future customers may overtake him. Therefore, we call this discipline *current priority discipline* and denote this model by CP.

We already showed that there may be numerous equilibria in this models. For example, always overtaking k customers, i.e., Σ_k, $k = 0, 1, 2 \ldots$, are equilibrium strategies. In particular strategy Σ_∞, in which an arriving customer overtakes all customers who are currently in the system. Note that Σ_∞ induces a last-come fist-served order of service.

Notice that C_o is a parameter that can be changed by a server, as opposed to C_w which is a given parameter. Denote by C_o^* the value which the server chooses in order to maximize its expected profit.

Theorem 6.1 *The maximum profit in the CP model among all equilibrium strategies is received from Σ_∞ with $C_o^* = \frac{C_w}{\mu-\lambda}$.*

Proof: $C_o = \frac{C_w}{\mu-\lambda}$ it is the maximum price for overtaking any customer and still Σ_∞ defines an equilibrium. In Theorem 5.1 we showed that under any other equilibrium strategy the price per each overtaken customer is not higher and the number of overtaken customers is not higher either. Therefore, the maximum profit is received from Σ_∞ with this price. ∎

We assume that the server can choose the equilibrium which maximizing its expected profit. Hence, it will choose Σ_∞ strategy in the CP model. The profit from a customer is the cost which this customer pays. In the CP model it is C_o per each overtaken customer. Denote by Π^{CP} the server's expected profit *per customer* in the CP model where an arriving customer always overtakes all present customers, i.e., Σ_∞ strategy, and the price per overtaken customer is the maximum price which satisfies the equilibrium conditions, i.e., C_o^*. The server's expected profit per customer (who overtakes all present customers) is $C_o^* L$, where $L = \frac{\lambda}{\mu-\lambda}$ is the expected number of customers in the system. Therefore,

$$\Pi^{\mathrm{CP}} = \frac{\lambda}{(\mu-\lambda)^2} C_w. \tag{30}$$

6.2 Maximum profit in the absolute priority (AP) model with threshold n=0

The second model which is analyzed by Adiri and Yechiali [1], and Hassin and Haviv [9] has two priority classes. In this model two FCFS queues are formed in front of a single server, one for priority customer and the other for ordinary customers. For a given threshold value $n \geqslant 0$, an arriving customer buys priority if and only if the number of customers in the ordinary queue is at least n. In other words, this is an absolute priority discipline, and therefore denote this model by AP. If a customer purchases priority then he overtakes all customers in the ordinary queue, and becomes the last customer in the priority queue. The price for becoming a lower priority ordinary customer is 0, this assumption is without loss of generality since there is no balking or reneging.

Denote by θ the price of purchasing priority, and by $W(n)$ the expected time in the system of the last customer in the ordinary queue when there are no customers in the priority queue and n in the ordinary one, and all use the pure threshold strategy n. The following theorem is proved by Hassin and Haviv [9]:

Theorem 6.2 *The integer threshold strategy* n, $n \geqslant 1$, *specifies an equilibrium if and only if* $\theta + \frac{C_w}{\mu} - \frac{C_w}{\mu-\lambda} \leqslant C_w W(n) \leqslant \theta + \frac{C_w}{\mu}$. *The threshold* $n = 0$ *specifies an equilibrium if and only if* $\theta + \frac{C_w}{\mu} \leqslant \frac{C_w}{\mu-\lambda}$.

The profit from a customer is the cost which this customer pays. In the AP model it is θ, if a customer buys priority, otherwise it is zero. Denote by $\Pi^{\mathrm{AP}}(n)$ the server's expected profit *per customer* in the AP model as a function of a threshold n, and by θ_{max} the maximum price for buying priority which satisfies the equilibrium conditions. In this section we compute the maximum profit in the AP model with threshold $n = 0$. In this case all customers use the pure threshold strategy $n = 0$, therefore the strategy is always buying priority, and from Theorem 6.2, $\theta_{max} = \frac{\lambda}{\mu(\mu-\lambda)} C_w$, so that $\Pi^{\mathrm{AP}}(0) = \theta_{max}$.

Since $\frac{\lambda}{\mu(\mu-\lambda)} C_w < \frac{\lambda}{(\mu-\lambda)^2} C_w$, it follows that, $\Pi^{\mathrm{AP}}(0) < \Pi^{\mathrm{CP}}$, i.e., the server's expected profit per arrival in the CP model is greater than the server's expected profit per arrival in the AP model with threshold $n = 0$.

6.3 Maximum profit in the absolute priority (AP) model with threshold n$\geqslant$ 1

In this section we compute the maximum profit in the AP model with threshold $n \geqslant 1$. Denote by P_n the probability that the number of customers in the system (both ordinary and priority queues) is at least n, in the AP model under the threshold strategy n. $P_n = P(L \geqslant n) = (\frac{\lambda}{\mu})^n$. We assume that all customers use the pure threshold strategy $n \geqslant 1$. From Theorem 6.2, in this case $\theta_{max} = C_w \left[W(n) + \frac{1}{\mu-\lambda} - \frac{1}{\mu}\right] = C_w \left[W(n) + \frac{\lambda}{\mu(\mu-\lambda)}\right]$. Since an arriving customer buys priority if and only if the number of customers in the queue is at least n, $\Pi^{\text{AP}}(n) = \theta_{max} P_n = \theta_{max} \left(\frac{\lambda}{\mu}\right)^n$, or equivalently

$$\Pi^{\text{AP}}(n) = C_w \left[W(n) + \frac{\lambda}{\mu(\mu-\lambda)}\right] \left(\frac{\lambda}{\mu}\right)^n . \tag{31}$$

We say that a customer in the ordinary queue is in state (i, j) if there are exactly i customers in front of him in the ordinary queue (including the one in service), exactly j customers behind him in the ordinary queue, and no customers in the priority queue. We denote by $f_{i,j}(n)$ the expected (residual) waiting time of a customer in state (i, j) given that all future customers adopt the pure threshold strategy n. In addition let $f_{-1,j}(n) = 0$. Hence $W(n) = f_{n-1,0}(n)$.

We now express the equations for calculating $W(n)$.

- Suppose that the state is (i, j) such that $i + j < n - 1$, $i = 1, \ldots, n-2$ and $j = 0, \ldots, n-2$. The expected time till the next arrival or service completion is $\frac{1}{\lambda+\mu}$. With probability $\frac{\mu}{\lambda+\mu}$ the service completion occurs before a new arrival, and then the customer's expected residual waiting time is $f_{i-1,j}(n)$. With probability $\frac{\lambda}{\lambda+\mu}$ a new customer arrives before a service completion occurs and since $i + j < n - 1$, i.e., there are less than n customers in the system, a new customer does not overtake any customer and the customer's expected residual waiting time is $f_{i,j+1}(n)$. Therefore, for i, j such that $i+j < n-1$, $i = 1, \ldots, n-2$ and $j = 0, \ldots, n-2$,

$$f_{i,j}(n) = \frac{1}{\lambda+\mu} + \frac{\mu}{\lambda+\mu} f_{i-1,j}(n) + \frac{\lambda}{\lambda+\mu} f_{i,j+1}(n). \tag{32}$$

- Suppose that the state is $(i, n-i-1)$, $i = 1, \ldots, n-1$, i.e., there are n customers in the system. Then, all future arrivals will overtake the customer, till the number of customers in the queue is reduced by one. This is a busy period. Hence, for $i = 0, \ldots, n-2$,

$$f_{i,n-i-1}(n) = \frac{1}{\mu-\lambda} + f_{i-1,n-i-1}(n). \tag{33}$$

In particular,

$$W(n) = f_{n-1,0}(n) = \frac{1}{\mu - \lambda} + f_{n-2,0}(n). \tag{34}$$

- If the state is $(0, j)$, $j \in 0, 1 \ldots, n-2$, the expected time till the next arrival or service completion is $\frac{1}{\lambda+\mu}$. With probability $\frac{\lambda}{\lambda+\mu}$ a new customer arrives before a service completion occurs and since $j < n-1$, i.e., there are less than n customers in the system, a new arriving customer does not overtake any customer and the customer's expected residual waiting time is $f_{0,j+1}(n)$. Therefore,

$$f_{0,j}(n) = \frac{1}{\lambda + \mu} + \frac{\lambda}{\lambda + \mu} f_{0,j+1}(n), \quad j \in 0, 1, \ldots, n-2. \tag{35}$$

Lemma 6.3

$W(1) = \frac{1}{\mu-\lambda}$ *and* $\Pi^{\text{AP}}(1) = C_w \frac{\lambda+\mu}{\mu(\mu-\lambda)} \frac{\lambda}{\mu}$*;*

$W(2) = \frac{2\mu+\lambda}{\mu^2-\lambda^2}$*, and* $\Pi^{\text{AP}}(2) = C_w \frac{2\mu^2+2\lambda\mu+\lambda^2}{\mu(\mu^2-\lambda^2)} (\frac{\lambda}{\mu})^2$*;*

$W(3) = \frac{3\mu^3+7\lambda\mu^2+4\lambda^2\mu+\lambda^3}{(\lambda+\mu)^2(\mu^2-\lambda^2)}$*, and* $\Pi^{\text{AP}}(3) = C_w \frac{3\mu^4+8\lambda\mu^3+7\lambda^2\mu^2+4\lambda^3\mu+\lambda^4}{\mu(\lambda+\mu)^2(\mu^2-\lambda^2)} (\frac{\lambda}{\mu})^3$*.*

In these cases $\Pi^{\text{AP}}(n) < \Pi^{\text{CP}}$*.*

Proof: If $n = 1$, then all new arrivals buy priority and overtake the present ordinary customer. In this case when the ordinary customer's service ends the system becomes empty. Thus his waiting time amounts to a busy period. Therefore,

$$W(1) = \frac{1}{\mu - \lambda}. \tag{36}$$

Substituting (36) in (31) gives $\Pi^{\text{AP}}(1) = C_w \left[\frac{1}{\mu-\lambda} + \frac{\lambda}{\mu(\mu-\lambda)}\right] \frac{\lambda}{\mu}$, or equivalently,

$$\Pi^{\text{AP}}(1) = C_w \frac{\lambda + \mu}{\mu(\mu - \lambda)} \frac{\lambda}{\mu}. \tag{37}$$

Comparing $\Pi^{\text{AP}}(1)$ from (37) to Π^{CP} from (30) we get that $\Pi^{\text{AP}}(1) < \Pi^{\text{CP}}$.

Observation 6.4 $W(2) = f_{1,0}(2)$.

Now we compute $W(2)$.

- Suppose a customer is in state (0,0). The expected time till the next arrival or service completion occurs is $\frac{1}{\lambda+\mu}$. With probability $\frac{\lambda}{\lambda+\mu}$ a new customer arrives before a service completion occurs. Then, the new arrival observes one customer upon arrival, therefore he does not buy priority and does not overtake the present customer in the ordinary queue. Hence,

$$f_{0,0}(2) = \frac{1}{\lambda+\mu} + \frac{\lambda}{\lambda+\mu} f_{0,1}(2). \tag{38}$$

- Suppose a customer is in state (0,1). All future arrivals will observe two customers or more upon arrival, therefore, they will buy priority and overtake the present customers in the ordinary queue till the number of customers in the ordinary queue is reduced by one, and it is equal to a busy period. Hence,

$$f_{0,1}(2) = \frac{1}{\mu-\lambda}. \tag{39}$$

- Suppose a customer is in state (1,0). All future arrivals will observe two customers or more upon arrival, therefore, they will buy priority and will overtake the present customers in the ordinary queue till the number of customers in the ordinary queue is reduced by one, and it is equal to a busy period. Hence,

$$f_{1,0}(2) = \frac{1}{\mu-\lambda} + f_{0,0}(2). \tag{40}$$

Substituting (39) in (38) gives,

$$f_{0,0}(2) = \frac{1}{\lambda+\mu} + \frac{\lambda}{\mu^2-\lambda^2} = \frac{\mu}{\mu^2-\lambda^2}. \tag{41}$$

Substituting (41) in (40) gives, $f_{1,0}(2) = \frac{\lambda+2\mu}{\mu^2-\lambda^2}$. Since $W(2) = f_{1,0}(2)$,

$$W(2) = \frac{\lambda+2\mu}{\mu^2-\lambda^2}. \tag{42}$$

Substituting (42) in (31) gives $\Pi^{\mathrm{AP}}(2) = C_w \left[\frac{\lambda+2\mu}{\mu^2-\lambda^2} + \frac{\lambda}{\mu(\mu-\lambda)}\right]\left(\frac{\lambda}{\mu}\right)^2$, or equivalently,

$$\Pi^{\mathrm{AP}}(2) = C_w \frac{2\mu^2 + 2\lambda\mu + \lambda^2}{\mu(\mu^2 - \lambda^2)} \left(\frac{\lambda}{\mu}\right)^2. \tag{43}$$

Comparing $\Pi^{\mathrm{AP}}(2)$ from (43) to Π^{CP} from (30) we get that $\Pi^{\mathrm{AP}}(2) < \Pi^{\mathrm{CP}}$.

Observation 6.5 *$W(3) = f(2,0)(3)$.*

Now we compute $W(3)$.

- Suppose a customer is in state (0,0). The expected time till the next arrival or service completion occurs is $\frac{1}{\lambda+\mu}$. With probability $\frac{\lambda}{\lambda+\mu}$ a new customer arrives before a service completion occurs. Then, the new arrival observes one customer upon arrival, therefore he does not buy priority and does not overtake the present customer in the ordinary queue. Hence,

$$f_{0,0}(3) = \frac{1}{\lambda+\mu} + \frac{\lambda}{\lambda+\mu} f_{0,1}(3). \tag{44}$$

- Suppose a customer is in state (0,1). The expected time till the next arrival or service completion occurs is $\frac{1}{\lambda+\mu}$. With probability $\frac{\lambda}{\lambda+\mu}$ a new customer arrives before a service completion occurs. Then, the new arrival observes two customer upon arrival, therefore he does not buy priority and does not overtake the present customer in the ordinary queue. Hence,

$$f_{0,1}(3) = \frac{1}{\lambda+\mu} + \frac{\lambda}{\lambda+\mu} f_{0,2}(3). \tag{45}$$

- Suppose a customer is in state (0,2). All future arrivals will observe three customers or more upon arrival, therefore, they will buy priority and will overtake the present customers in the ordinary queue till the number of customers in the ordinary queue is reduced by one, and it is equal to a busy period. Hence,

$$f_{0,2}(3) = \frac{1}{\mu - \lambda}. \tag{46}$$

- Suppose a customer is in state (1,0). The expected time till the next arrival or service completion occurs is $\frac{1}{\lambda+\mu}$. With probability $\frac{\lambda}{\lambda+\mu}$ a new customer arrives before a service completion occurs. Then, the new arrival observes two customer upon arrival, therefore he does not buy priority and does not overtake the present customer in the ordinary queue. With probability $\frac{\mu}{\lambda+\mu}$ a service completion occurs before an arrival of a new customer, then the customer's expected time is $f_{0,0}(3)$. Hence,

$$f_{1,0}(3) = \frac{1}{\lambda+\mu} + \frac{\mu}{\lambda+\mu} f_{0,0}(3) + \frac{\lambda}{\lambda+\mu} f_{1,1}(3). \tag{47}$$

- Suppose a customer is in state (1,1). All future arrivals will observe three customers or more upon arrival, therefore, they will buy priority and overtake the present customers in the ordinary queue till, the number of customers in the ordinary queue is reduced by one, and it is equal to a busy period. Hence,

$$f_{1,1}(3) = \frac{1}{\mu-\lambda} + f_{0,1}(3). \tag{48}$$

- Suppose a customer is in state (2,0). All future arrivals will observe three customers or more upon arrival, therefore, they will buy priority and will overtake the present customers in the ordinary queue till the number of customers in the ordinary queue is reduced by one, and it is equal to a busy period. Hence,

$$f_{2,0}(3) = \frac{1}{\mu-\lambda} + f_{1,0}(3). \tag{49}$$

Substituting (46) in (45), gives

$$f_{0,1}(3) = \frac{\mu}{\mu^2-\lambda^2}. \tag{50}$$

Substituting (50) in (48), gives

$$f_{1,1}(3) = \frac{2\mu+\lambda}{\mu^2-\lambda^2}. \tag{51}$$

Substituting (50) in (44), gives

$$f_{0,0}(3) = \frac{\mu^2 - \lambda^2 + \lambda\mu}{(\lambda+\mu)(\mu^2-\lambda^2)}. \tag{52}$$

Substituting (52) and (51) in (47) and the result in (49) gives, $f_{2,0}(3) = \frac{3\mu^3+7\lambda\mu^2+4\lambda^2\mu+\lambda^3}{(\lambda+\mu)^2(\mu^2-\lambda^2)}$.

Since $W(3) = f_{2,0}$,

$$W(3) = \frac{3\mu^3 + 7\lambda\mu^2 + 4\lambda^2\mu + \lambda^3}{(\lambda+\mu)^2(\mu^2-\lambda^2)}. \tag{53}$$

Substituting (53) in (31) gives $\Pi^{\mathrm{AP}}(3) = C_w \left[\frac{3\mu^3+7\lambda\mu^2+4\lambda^2\mu+\lambda^3}{(\lambda+\mu)^2(\mu^2-\lambda^2)} + \frac{\lambda}{\mu(\mu-\lambda)}\right]\left(\frac{\lambda}{\mu}\right)^3$, or equivalently,

$$\Pi^{\mathrm{AP}}(3) = C_w \frac{3\mu^4 + 8\lambda\mu^3 + 7\lambda^2\mu^2 + 4\lambda^3\mu + \lambda^4}{\mu(\lambda+\mu)^2(\mu^2-\lambda^2)}\left(\frac{\lambda}{\mu}\right)^3. \tag{54}$$

Comparing $\Pi^{\mathrm{AP}}(3)$ from (54) to Π^{CP} from (30) we get that $\Pi^{\mathrm{AP}}(3) < \Pi^{\mathrm{CP}}$.

■

Since it is difficult to find general expressions to $f_{i,j}(n)$, we numerically compute these values. In all cases we found that $\Pi^{\mathrm{AP}}(n) < \Pi^{\mathrm{CP}}$. Some results are illustrated in the next Subsection 6.4.

6.4 Numerical Analysis of profit maximization

The graphs in Figure 4 present the server's expected profit per customer in the AP model as a function of the threshold n and arrival rate λ. For every λ the server's expected profit is higher when the threshold n is smaller. There are λ values for which the function is convex, for example $\lambda = 0.3$. There are λ values for which the function is concave, for example $\lambda = 0.99$, and there are λ values for which the function is neither convex nor concave, for example $\lambda = 0.9$.

As presented in Tables 2 and 3, Π^{CP} is much grater than $\Pi^{\mathrm{AP}}(n)$ for all presented parameters. Therefore, the server can obtain a higher profit in our model.

In addition, we see in Figure 5, as expected, that when the service is slower, i.e., μ is lower, the server's profit is higher. This result is expected since there is no balking.

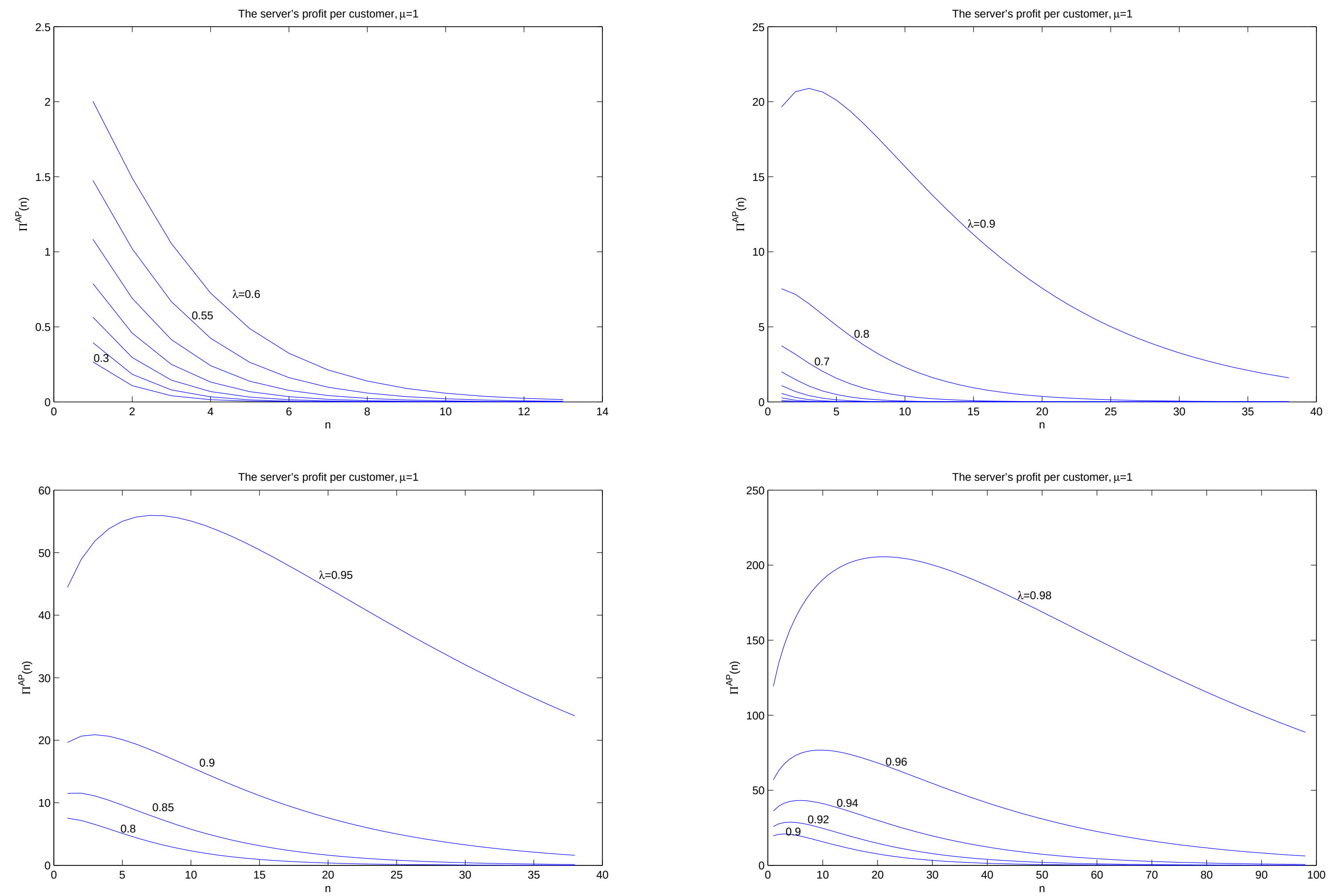

Figure 4: The server's expected profit $\Pi^{\text{AP}}(n)$ as a function of the threshold n and arrival rate λ.

λ	$\Pi^{\mathrm{AP}}(4)$	$\Pi^{\mathrm{AP}}(7)$	$\Pi^{\mathrm{AP}}(10)$	Π^{CP}
0.1	0.0004	0.0000	0.0000	0.0123
0.2	0.0073	0.0001	0.0000	0.0625
0.3	0.0406	0.0018	0.0001	0.1837
0.4	0.1452	0.0143	0.0012	0.4444
0.5	0.4149	0.0768	0.0126	1.0000
0.6	1.0556	0.3241	0.0895	2.2500
0.7	2.5746	1.2067	0.5149	5.4444
0.8	6.5393	4.4083	2.7315	16.0000
0.9	20.8894	19.3777	16.6511	81.0000

Table 2: Server's expected profit per customer in CP and AP models, $\mu = 1$.

μ	$\Pi^{\mathrm{AP}}(4)$	$\Pi^{\mathrm{AP}}(7)$	$\Pi^{\mathrm{AP}}(10)$	Π^{CP}
0.2	2.0743	0.3838	0.0630	5.0000
0.3	0.2143	0.0126	0.0006	0.8333
0.4	0.0466	0.0012	0.0000	0.2778
0.5	0.0146	0.0002	0.0000	0.1250
0.6	0.0057	0.0000	0.0000	0.0667
0.7	0.0026	0.0000	0.0000	0.0397
0.8	0.0013	0.0000	0.0000	0.0255
0.9	0.0007	0.0000	0.0000	0.0174
1.0	0.0004	0.0000	0.0000	0.0123

Table 3: Server's expected profit per customer in CP and AP models, $\lambda = 0.1$.

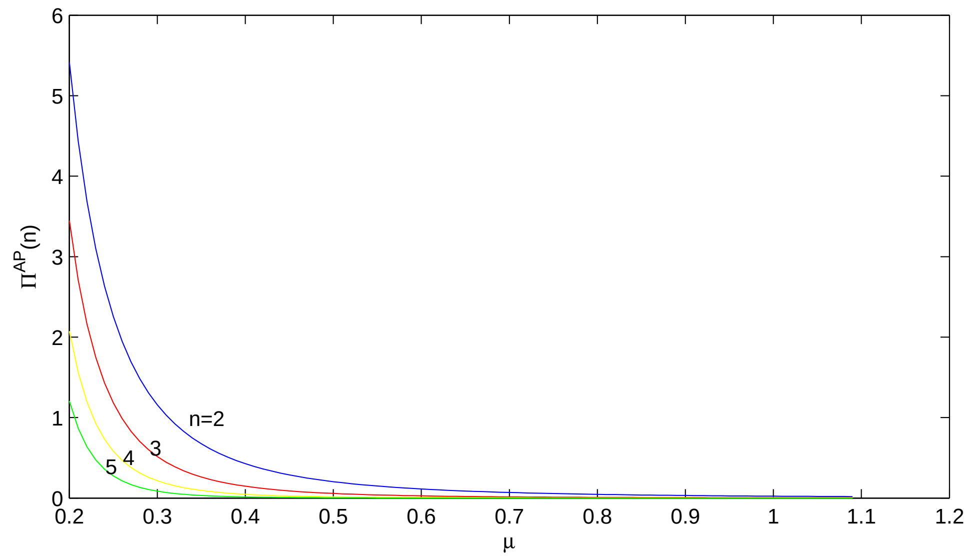

Figure 5: The server's expected profit per customer in AP model as a function of μ.

7 A threshold strategy for overtaking a single customer

7.1 Overtaking one customer - pure threshold strategy

The pure threshold strategy σ_n is defined as follows: a new customer overtakes one customer if there are n or more customers in the system, and does not overtake any customer otherwise.

Theorem 7.1 *The pure threshold strategy σ_n defines an equilibrium if and only if $\frac{1}{\mu-\lambda} \leqslant \frac{C_o}{C_w} \leqslant \frac{\mu+\lambda}{\mu(\mu-\lambda)}$.*

Proof: When all others apply the pure threshold strategy σ_n, a new customer's best response does not overtake more than one customer, since by overtaking one customer the new customer guarantees his place in the queue and from the assumption of our model that $\frac{C_w}{\mu} < C_o$ there is no benefit in overtaking more than one. In addition, if a customer observes $n - j$ customers, $j = 2, 3, \ldots, n$ then not overtaking any customer is the best response since he will never be overtaken, and from the assumption of our model that $\frac{C_w}{\mu} < C_o$ there is no benefit in overtaking any customer.

- Suppose that a new customer observes $n-1$ customers. If he does not overtake, all future arrivals will overtake him as long as his position is n or more. The time it takes for the new customer till he reaches position $n-1$ is equal to a busy period. So the customer's expected cost is $\frac{C_w}{\mu}\left(\frac{1}{1-\rho}+n-1\right)$. Otherwise, if the new customer overtakes a single customer, he guarantees his place in the queue and his expected cost is $C_w\frac{n-1}{\mu}+C_o$. In symmetric equilibrium the best response of a new customer is not overtaking. Hence, $\frac{C_w}{\mu}(\frac{1}{1-\rho}+n-1) \leqslant C_w\frac{n-1}{\mu}+C_o$, and this inequality gives the first condition for an equilibrium, $\frac{1}{\mu-\lambda} \leqslant \frac{C_o}{C_w}$.

- Suppose that a new customer observes $n+j$ customers, $j=0,1,2,\ldots$, and doesn't overtake any customer. Then all future arrivals will overtake him as long as his position is n or more. Hence his waiting time consists of $j+2$ busy periods (reducing the number of customers by $j+2$), plus $n-1$ service periods. His expected cost is then $\frac{C_w}{\mu}\left(\frac{j+2}{1-\rho}+n-1\right)$. Otherwise, if a new customer overtakes a single customer, he guarantees his place in the queue and his expected cost is $C_w\frac{n+j}{\mu}+C_o$. In symmetric equilibrium the best response of a new customer is overtaking. Hence, $C_w\frac{n+j}{\mu}+C_o \leqslant \frac{C_w}{\mu}\left(\frac{j+2}{1-\rho}+n-1\right)$, or equivalently $\frac{C_o}{C_w} \leqslant \frac{\mu+\lambda+j\lambda}{\mu(\mu-\lambda)}$, and $\frac{\mu+\lambda+j\lambda}{\mu(\mu-\lambda)}$ is minimum for $j=0$. Therefore, $\frac{C_o}{C_w} \leqslant \frac{\mu+\lambda}{\mu(\mu-\lambda)}$ is the second condition for an equilibrium.

■

7.2 Overtaking one customer - mixed threshold strategy

The mixed threshold strategy $\sigma_{n,p}$ where $0 < p < 1$, is defined as follows: a new customer overtakes one customer if there are at least $n+1$ customers in the system, does not overtake any customer if there are at most $n-1$ customers in the system, and if there are exactly n customers in the system he overtakes one customer with probability p, and does not overtake any customer otherwise.

Theorem 7.2 *The mixed threshold strategy $\sigma_{n,p}$ defines an equilibrium if and only if $\frac{1}{\mu-\lambda} \leqslant \frac{C_o}{C_w} \leqslant \frac{\mu+\lambda}{\mu(\mu-\lambda)}$ and $p = p_e$, where $p_e = \frac{(\mu+\lambda)(C_o(\mu-\lambda)-C_w)}{C_o\lambda(\mu-\lambda)}$.*

Proof: Define $f_i(p)$ to be the expected waiting time in position i, when i is the last customer in the queue, given that all customers follow the strategy $\sigma_{n,p}$.

The expected waiting time in position n is computed as follows: the expected time till either a service completion or a new arrival occurs is $\frac{1}{\lambda+\mu}$. With probability $\frac{\mu}{\lambda+\mu}$ the service completion occurs before a new customer arrives, in this position his place is guaranteed, and the expected waiting time $f_{n-1}(p)$, consists of $n-1$ service periods. With probability $\frac{\lambda}{\lambda+\mu}$ a new customer arrives before a service completion occurs, then the new arrival overtakes the n-th customer with probability p and his expected waiting time is $f_{n+1}(p)$. Otherwise, if the new customer doesn't overtake the n-th customer, the n-th customer's position is guaranteed, and his expected waiting time is n service periods. The expected waiting time in position $n+1$ is computed as follows: in this position all future arrivals will overtake the last customer, therefore his waiting time will be the busy period (which means to reduce the number of customers in the system to n), plus $f_n(p)$. Hence, we get

$$f_n(p) = \frac{1}{\mu+\lambda} + \frac{\mu}{\mu+\lambda}\frac{n-1}{\mu} + \frac{\lambda}{\mu+\lambda}\left[p f_{n+1}(p) + (1-p)\frac{n}{\mu}\right]. \tag{55a}$$

$$f_{n+1}(p) = \frac{1}{\mu}\frac{1}{1-\rho} + f_n(p) = \frac{1}{\mu-\lambda} + f_n(p). \tag{55b}$$

Substituting (55b) in (55a) gives

$$f_n(p) = \frac{1}{\mu-\lambda-\lambda p}\left[n + \frac{\lambda p}{\mu-\lambda} + (1-p)\frac{\lambda n}{\mu}\right]. \tag{56}$$

Under the pure threshold strategy $\sigma_{n,p}$ the new customer does not overtake more than one customer, since by overtaking one customer the new customer guarantees his place in the queue and from the assumption of our model that $\frac{C_w}{\mu} < C_o$ there is no benefit in overtaking more than one.

- Suppose that a new customer observes n customers. If he overtakes a single customer, he guarantees his place in the queue and his expected waiting time is n service periods, plus C_o. Otherwise, if he does not overtake, all future arrivals will overtake him, therefore his expected waiting time is a single busy period, plus $f_n(p)$. In equilibrium a new customer is indifferent between overtaking a single customer or not overtaking. Hence, $C_w f_{n+1}(p) = C_w \frac{n}{\mu} + C_o$, or $C_w \left[\frac{1}{\mu-\lambda} + f_n(p)\right] = C_w \frac{n}{\mu} + C_o$. Substituting $f_n(p)$ from (56) gives $p_e = \frac{(\mu+\lambda)(C_o(\mu-\lambda)-C_w)}{C_o\lambda(\mu-\lambda)}$.

- Because p_e is a probability, we require that $0 < p_e < 1$. The denominator of p_e is always positive, so the numerator must be positive too. Therefore $C_o(\mu-\lambda) - C_w > 0$, or $\frac{C_o}{C_w} > \frac{1}{\mu-\lambda}$, and this is one of the conditions for an equilibrium in a pure threshold strategy.

  The condition for $p_e < 1$ is $(\mu+\lambda)(C_o(\mu-\lambda) - C_w) < C_o\lambda(\mu-\lambda)$, or $\frac{C_o}{C_w} < \frac{\mu+\lambda}{\mu(\mu-\lambda)}$, and this is the additional condition for an equilibrium in a pure threshold strategy.

- If p_e is an equilibrium strategy, then the best response of a new customer who observes $n-1$ customer is not to overtake. If he does not overtake, he will be n in the system, and his expected waiting time will be $f_n(p)$. Otherwise, if he overtakes, he guarantees his place in the queue and his expected waiting time is $n-1$ service periods, plus C_o. Therefore we should get $C_w f_n(p) < C_w \frac{n-1}{\mu} + C_o$, or $\frac{C_w\lambda}{C_o(\mu-\lambda)} > 0$, and this is always true.

■

8 Concluding remarks

- We did not analyze social optimization since the customers are statistically identical, i.e., we assume homogeneous time values, and in addition we do not allow balking. Hassin [7] observed that if there is an option for balking then priorities have a positive influence on social welfare even if all customers are identical. If the customers are not identical, for example, have heterogeneous time values then apparently in our model the customers with lower time values will be at the end of the queue and others will overtake them. These issues are for further study.

- When assuming heterogeneous time values, the simplest and most intuitive model of an incentive compatible pricing scheme is that of Ghanem [6]. Ghanem proved the intuitive result that for social optimization higher priority should be given to customers with a higher time value.

- Fairness among customers is a fundamental issue for queueing systems. In many situations we notice that customers wish for fair service and fair waiting time. The issue of fairness is raised frequently in the context of evaluating queueing policies and its resolution is not simple at all. Avi-Itzhak and Levy [2] propose a fairness measure enabling to quantitatively measure and compare the level of fairness associated with various queueing systems. They propose yardsticks that can be used as standards for evaluating the fairness of various queueing systems with one class of customers, and for comparing different disciplines to each other. They focused on the issue of customer seniority which is crucial in many queueing systems and used an axiomatic approach to develop fairness measure that is based on this notion. Raz, Avi-Itzhak, and Levy [15] develop a quantitative model for studying priority and classification systems, focusing on the relative fairness of these mechanism. Their analysis provides a measure of fairness for these systems. They limit the discussion to systems where job classification is based only on service characteristics. One of their results is that from fairness perspective, providing preferential service to shorter jobs may be justified in many cases. For comparison, they provide the fairness analysis for an equivalent system where jobs are served in the order of arrivals (FCFS). They conclude that in many cases prioritization of short jobs over the long jobs leads to higher fairness (than that of FCFS), nonetheless, in some cases FCFS is more fair. In queueing systems with priorities the issue is how priorities and preferential service affect fairness. In queueing systems with priorities which involve costs (waiting cost, priority cost) such as our model the issue of how priorities and preferential service affect fairness has not been explored and evaluated at all. This is an interesting subject for future research.

Appendix-A

In Section 4 we analyzed pure strategies in the observable model where customers observe the number of people who are already in the system, and then decide how many customers to overtake. Those pure strategies are defined by a vector $(k_1, k_2, \ldots)$, where k_i is the the number of customers that a new customer overtakes if he observes i customers in the system upon arrival. Clearly, $k_i \leqslant i$. We found some very surprising results of unexpected strategies which for some parameters are equilibrium strategies. In this appendix we give $(0,2,0,0,5,5)$ strategy as an example for such unexpected strategy.

The equations for $(0,2,0,0,5,5)$ strategy are:

$f_{i,j} = \frac{i+1}{j}, \quad \forall j \geqslant 5$

$f_{0,4} = \frac{1}{\lambda+\mu} + \frac{\lambda}{\lambda+\mu} f_{1,4}$

$f_{1,4} = \frac{1}{\lambda+\mu} + \frac{\mu}{\lambda+\mu} f_{0,4} + \frac{\lambda}{\lambda+\mu} f_{2,4}$

$f_{i,4} = \frac{1}{\lambda+\mu} + \frac{\mu}{\lambda+\mu} f_{i-1,4} + \frac{\lambda}{\lambda+\mu} \frac{i+1}{\mu}, \quad \forall i \geqslant 2$

$f_{0,3} = \frac{1}{\lambda+\mu} + \frac{\lambda}{\lambda+\mu} f_{0,4}$

$f_{1,3} = \frac{1}{\lambda+\mu} + \frac{\mu}{\lambda+\mu} f_{0,3} + \frac{\lambda}{\lambda+\mu} f_{2,3}$

$f_{2,3} = \frac{1}{\lambda+\mu} + \frac{\mu}{\lambda+\mu} f_{1,3} + \frac{\lambda}{\lambda+\mu} f_{3,3}$

$f_{i,3} = \frac{1}{\lambda+\mu} + \frac{\mu}{\lambda+\mu} f_{i-1,3} + \frac{\lambda}{\lambda+\mu} f_{i,4}, \quad \forall i \geqslant 3$

$f_{0,2} = \frac{1}{\lambda+\mu} + \frac{\lambda}{\lambda+\mu} f_{0,3}$

$f_{1,2} = \frac{1}{\lambda+\mu} + \frac{\mu}{\lambda+\mu} f_{0,2} + \frac{\lambda}{\lambda+\mu} f_{1,3}$

$f_{2,2} = \frac{1}{\lambda+\mu} + \frac{\mu}{\lambda+\mu} f_{1,2} + \frac{\lambda}{\lambda+\mu} f_{3,2}$

$f_{3,2} = \frac{1}{\lambda+\mu} + \frac{\mu}{\lambda+\mu} f_{2,2} + \frac{\lambda}{\lambda+\mu} f_{4,2}$

$f_{i,2} = \frac{1}{\lambda+\mu} + \frac{\mu}{\lambda+\mu} f_{i-1,2} + \frac{\lambda}{\lambda+\mu} f_{i,3}, \quad \forall i \geqslant 4$

$f_{0,1} = \frac{1}{\lambda+\mu} + \frac{\lambda}{\lambda+\mu} f_{1,1}$

$f_{1,1} = \frac{1}{\lambda+\mu} + \frac{\mu}{\lambda+\mu} f_{0,1} + \frac{\lambda}{\mu+\lambda} f_{1,2}$

$f_{2,1} = \frac{1}{\lambda+\mu} + \frac{\mu}{\lambda+\mu} f_{1,1} + \frac{\lambda}{\lambda+\mu} f_{2,2}$

$f_{3,1} = \frac{1}{\lambda+\mu} + \frac{\mu}{\lambda+\mu} f_{2,1} + \frac{\lambda}{\lambda+\mu} f_{4,1}$

$f_{4,1} = \frac{1}{\lambda+\mu} + \frac{\mu}{\lambda+\mu} f_{3,1} + \frac{\lambda}{\lambda+\mu} f_{5,1}$

$f_{i,1} = \frac{1}{\lambda+\mu} + \frac{\mu}{\lambda+\mu} f_{i-1,1} + \frac{\lambda}{\lambda+\mu} f_{i,2}, \quad \forall i \geqslant 5$

$f_{0,0} = \frac{1}{\lambda+\mu} + \frac{\lambda}{\lambda+\mu} f_{0,1}$

$f_{1,0} = \frac{1}{\lambda+\mu} + \frac{\mu}{\lambda+\mu} f_{0,0} + \frac{\lambda}{\lambda+\mu} f_{2,0}$

$f_{2,0} = \frac{1}{\lambda+\mu} + \frac{\mu}{\lambda+\mu} f_{1,0} + \frac{\lambda}{\lambda+\mu} f_{2,1}$

$f_{3,0} = \frac{1}{\lambda+\mu} + \frac{\mu}{\lambda+\mu} f_{2,0} + \frac{\lambda}{\lambda+\mu} f_{3,1}$

$f_{4,0} = \frac{1}{\lambda+\mu} + \frac{\mu}{\lambda+\mu} f_{3,0} + \frac{\lambda}{\lambda+\mu} f_{5,0}$

$f_{5,0} = \frac{1}{\lambda+\mu} + \frac{\mu}{\lambda+\mu} f_{4,0} + \frac{\lambda}{\lambda+\mu} f_{6,0}$

$f_{i,0} = \frac{1}{\lambda+\mu} + \frac{\mu}{\lambda+\mu} f_{i-1,0} + \frac{\lambda}{\lambda+\mu} f_{i,1}, \quad \forall i \geqslant 6$

The equilibrium conditions are:

$\underline{k_1 = 0}:$

$\frac{C_o}{C_w} > f_{1,0} - f_{0,1}$

$\underline{k_2 = 2}:$

$\frac{C_o}{C_w} < f_{1,1} - f_{0,2}$

$\frac{C_o}{C_w} < \frac{1}{2}(f_{2,0} - f_{0,2})$

$\underline{k_3 = 0}:$

$\frac{C_o}{C_w} > \frac{1}{i}(f_{3,0} - f_{3-i,i}), \quad 1 \leqslant i \leqslant 3$

$\underline{k_4 = 0}:$

$\frac{C_o}{C_w} > \frac{1}{i}(f_{4,0} - f_{4-i,i}), \quad 1 \leqslant i \leqslant 4$

$\underline{k_5 = 5}:$

$\frac{C_o}{C_w} < \frac{1}{i}(f_{i,5-i} - f_{0,5}), \quad 1 \leqslant i \leqslant 5$

$\underline{k_6 = 5}:$

$\frac{C_o}{C_w} < \frac{1}{5}(f_{6,0} - f_{1,5})$

$\frac{C_o}{C_w} < \frac{1}{4}(f_{5,1} - f_{1,5})$

$\frac{C_o}{C_w} < \frac{1}{3}(f_{4,2} - f_{1,5})$

$\frac{C_o}{C_w} < \frac{1}{2}(f_{3,3} - f_{1,5})$

$\frac{C_o}{C_w} < f_{2,4} - f_{1,5}$

$\frac{C_o}{C_w} > f_{1,5} - f_{1,5}$

$\underline{k_j = 0}, \quad \forall j \geqslant 7:$

$\frac{C_o}{C_w} > \frac{1}{i}(f_{j,0} - f_{j-i,i}), \quad 1 \leqslant i \leqslant j$

For example $f_{i,j}$ matrix for $\lambda = 1, \mu = 1.4, \frac{C_o}{C_w} = 1$:

0.9232	1.2157	0.8323	0.9975	1.3940	0.7143	0.7143	0.7143	0.7143	0.7143
2.1594	1.9176	1.9003	2.3955	2.3456	1.4286	1.4286	1.4286	1.4286	1.4286
2.8902	2.9133	3.3072	3.3528	2.6778	2.1429	2.1429	2.1429	2.1429	2.1429
3.8976	4.3079	4.2768	3.6930	3.1692	2.8571	2.8571	2.8571	2.8571	2.8571
5.2850	5.2605	4.6343	4.1348	3.7534	3.5714	3.5714	3.5714	3.5714	3.5714
6.2274	5.5941	5.0611	4.6586	4.3919	4.2857	4.2857	4.2857	4.2857	4.2857
6.5468	5.9939	5.5537	5.2433	5.0619	5.0000	5.0000	5.0000	5.0000	5.0000
6.9256	6.4559	6.1027	5.8713	5.7504	5.7143	5.7143	5.7143	5.7143	5.7143
7.3620	6.9730	6.6970	6.5289	6.4496	6.4286	6.4286	6.4286	6.4286	6.4286
7.8515	7.5367	7.3259	7.2065	7.1552	7.1429	7.1429	7.1429	7.1429	7.1429
8.3877	8.1384	7.9807	7.8973	7.8643	7.8571	7.8571	7.8571	7.8571	7.8571
8.9636	8.7699	8.6540	8.5966	8.5756	8.5714	8.5714	8.5714	8.5714	8.5714
9.5722	9.4243	9.3404	9.3014	9.2882	9.2857	9.2857	9.2857	9.2857	9.2857

Figure 6: $f_{i,j}$ matrix for $\lambda = 1, \mu = 1.4, \frac{C_o}{C_w} = 1$

Appendix-B

In Section 5.3 we analyzed a mixed strategy $\Sigma_{k,\mathbf{p}}$ where customers are indifferent between overtaking k, $k-1$ or $k-2$ customers. We gave necessary and sufficient conditions to this strategy to define an equilibrium. In this appendix we give an example to show that the equilibrium conditions are not an empty range. Suppose that $\lambda = 1$, $\mu = 2$, $\frac{C_o}{C_w} = 0.7$, then $x = \frac{1}{2}, y = \frac{1}{2}, z = \frac{1}{4}, p_{k+1}^{k} = 0.7916, p_{k+1}^{k-1} = 0.0477$, for every $k \geqslant 1$ satisfy the all conditions.

It is meaning that if a customer observes $k-1$ customers upon arrival, then with probability $\frac{1}{2}$ he overtakes $k-2$ customers or all of them. If a customer observes k customers upon arrival, then with probability $\frac{1}{4}$ he overtakes $k-2$ or $k-1$ customers, and with probability $\frac{1}{2}$ he overtakes k customers. If a customer observes $k+1$ customers upon arrival, then with probability 0.1607 he overtakes $k-2$ customers, with probability 0.0477 he overtakes $k-1$ customers, and with probability 0.7916 he overtakes k customers. Finally, if a customer observes $k+2$ or more customers upon arrival, then with probability 0.625 he overtakes $k-2$ customers, and with probability 0.375 he overtakes k customers.

The equilibrium conditions are:

$f_{j,k-1} = \frac{C_o}{C_w} + \frac{j}{\mu}, j \geqslant 1,$

$f_{j,k-2} = 2\frac{C_o}{C_w} + \frac{j-1}{\mu}, j \geqslant 2,$

$f_{0,k-1} = f_{1,k-2} - \frac{C_o}{C_w}.$

For example lets take $k = 3$, and suppose that if a new customer observes 15 or more customers upon arrival he does not overtake any of them.

$p_2^1 = 0.5$, $p_2^2 = 0.5$,

$p_3^1 = 0.25$, $p_3^2 = 0.25$, $p_3^3 = 0.5$,

$p_4^1 = 0.1607$, $p_4^2 = 0.047$, $p_4^3 = 0.7916$,

$p_j^1 = 0.0.625$, $p_j^2 = 0$, $p_j^3 = 0.375$, $\forall j \geqslant 5$.

Calculating $f_{i,j}$:

j=3:

$f_{i,j} = \frac{i+1}{\mu}$, $\forall i \geqslant 0$

$f_{0,2} = \frac{1}{\lambda+\mu} + \frac{\lambda}{\lambda+\mu}\left((p_3^1 + p_3^2)f_{0,3} + p_3^3 f_{1,2}\right)$

$f_{1,2} = \frac{1}{\lambda+\mu} + \frac{\mu}{\lambda+\mu}f_{0,2} + \frac{\lambda}{\lambda+\mu}\left((p_4^1 + p_4^2)f_{1,3} + p_4^3 f_{2,2}\right)$

$f_{2,2} = \frac{1}{\lambda+\mu} + \frac{\mu}{\lambda+\mu}f_{1,2} + \frac{\lambda}{\lambda+\mu}\left((p_5^1 + p_5^2)f_{2,3} + p_5^3 f_{3,2}\right)$

...

$f_{11,2} = \frac{1}{\lambda+\mu} + \frac{\mu}{\lambda+\mu}f_{10,2} + \frac{\lambda}{\lambda+\mu}\left((p_{14}^1 + p_{14}^2)f_{11,3} + p_{14}^3 f_{12,2}\right)$

$f_{12,2} = \frac{1}{\lambda+\mu} + \frac{\mu}{\lambda+\mu}f_{11,2} + \frac{\lambda}{\lambda+\mu}\frac{13}{\mu}$

$f_{13,2} = \frac{1}{\lambda+\mu} + \frac{\mu}{\lambda+\mu}f_{12,2} + \frac{\lambda}{\lambda+\mu}\frac{14}{\mu}$

$f_{0,1} = \frac{1}{\lambda+\mu} + \frac{\lambda}{\lambda+\mu}\left(p_2^1 f_{0,2} + p_2^2 f_{1,1}\right)$

$f_{1,1} = \frac{1}{\lambda+\mu} + \frac{\mu}{\lambda+\mu}f_{0,1} + \frac{\lambda}{\lambda+\mu}\left(p_3^1 f_{1,2} + (p_3^2 + p_3^3)f_{2,1}\right)$

$f_{2,1} = \frac{1}{\lambda+\mu} + \frac{\mu}{\lambda+\mu}f_{1,1} + \frac{\lambda}{\lambda+\mu}\left(p_4^1 f_{2,2} + (p_4^2 + p_4^3)f_{3,1}\right)$

...

$f_{12,1} = \frac{1}{\lambda+\mu} + \frac{\mu}{\lambda+\mu}f_{11,1} + \frac{\lambda}{\lambda+\mu}\left(p_{14}^1 f_{12,2} + (p_{14}^2 + p_{14}^3)f_{13,1}\right)$

$f_{13,1} = \frac{1}{\lambda+\mu} + \frac{\mu}{\lambda+\mu}f_{12,1} + \frac{\lambda}{\lambda+\mu}f_{13,2}$

0.6478	0.6333	0.5000
1.3405	1.2996	1.0000
1.9129	1.7997	1.5000
2.4136	2.2998	2.0000
2.9220	2.7999	2.5000
3.4293	3.2999	3.0000
3.9348	3.7999	3.5000
4.4388	4.2999	4.0000
4.9418	4.8000	4.5000
5.4439	5.3000	5.0000
5.9455	5.7999	5.5000
6.4463	6.2997	6.0000
6.9446	6.7979	6.5000
7.4282	7.2848	7.0000
7.8171	7.6899	7.5000

Figure 7: $f_{i,j}$ matrix $0 \leqslant i \leqslant 13$ and $j = 1, 2, 3$. $\lambda = 1$, $\mu = 2$, $\frac{C_o}{C_w} = 0.7$

References

[1] Adiri, I. and U.Yechiali (1974), "Optimal priority purchasing and pricing decisions in nonmonpoly and monopoly queues," *Operations Research* **22**, 1051-1066.

[2] Avi-Itzhak, B. and H. Levy (2004), "On measuring fairness in queues," *Advances in Applied Probability* **36**, 919-936.

[3] Conway, R.W., W.L. Maxwell, and L. W. Miller, *Theory of scheduling*, Addison-Wesley Pub. Co., Reading Mass, 1967.

[4] Edelson, N.M. and K. Hildebrand (1975), "Congestion tolls for Poisson queueing process," *Econometrica* **43**, 81-92.

[5] Glazer, A. and R. Hassin (1986), "Stable priority purchasing in queues," *Operations Research Letters* **4**, 285-288.

[6] Ghanem, S.B. (1975), "Computing central optimization by a pricing priority policy," *IBM Systems Journal* **14**, 272-292.

[7] Hassin, R. (1985), "On the optimality of first-come last-served queues," *Econometrica* **53**, 201-202.

[8] Hassin, R. (1995), "Decentralized regulation of a queue," *Management Science* **41**, 163-173.

[9] Hassin, R. and M. Haviv (1997), "Equilibrium threshold strategies: the case of queues with priorities," *Operations Research* **45**, 966-973.

[10] Hassin, R., J. Puerto, and F. R. Fernandez (2009), "The use of relative priorities in optimizing the performance of a queueing system," *European Journal of Operations Research* **193**, 476-483.

[11] Kleinrock, L., *Queueing Systems* (Vol. 1: Theory, Vol.2: Computer Applications), Wiley, 1975 and 1976.

[12] Kleinrock, L. (1967), "Optimal Bribing for Queuing Position," *Operations Research* **15**, 304-318.

[13] Lui, F.T (1985), "An equilibrium queuing model of bribery," *Journal of Political Economy* **93**, 760-781.

[14] Myrdal, G. (1968), *Asian Drama: An Inquiry into the Poverty of Nations*, Pantheon, New York.

[15] Raz, D., B. Avi-Itzhak and H. Levy (2004), "Classes, priorities and fairness in queueing systems," *Technical Report* RRR-21-2004, RUTCOR, Rutgers University.

[16] Rosenblum, D.M. (1992), "Allocation of waiting time by trading in position on a G/M/s queue," *Operations Research* **40**, 338-342.