National and Kapodistrian University of Athens

Department of Mathematics

The influence of information time in the strategic customer behavior

Thesis submitted to the National and Kapodistrian University of Athens, Department of Mathematics in partial fulfillment of the requirements for the degree MSc in Statistics and Operations Research

Anastas Baci

Supervisor: Antonis Economou

Athens, 2018

Approved on .... by the Evaluation Committee composed of:

Name-Surname	Academic rank	Signature
Antonis Economou (supervisor)	Professor	....
Apostolos Burnetas	Professor	....
Demetrios Fakinos	Associate Professor	....

Acknowledgments

I would like to thank my MSc advisor, professor Antonis Economou for the guidance, encouragement and useful advice he has provided. Without him this thesis would not have been possible. I would also like to thanks professor Apostolos Burnetas, and associate professor Demetrios Fakinos for accepting to be members of the evaluation committee for this thesis.

Contents

Chapter 1

Introduction

1.1 Queueing models

In general we do not like to wait. But reduction of waiting time usually requires extra investments. To decide whether or not to invest, it is important to know the effect of the investment on the waiting time, so we need models to analyze such situations.

For example, consider a telephone system whose function is to provide communication paths between pairs of telephone sets (customers) on demand. The provision of a permanent communication path between each pair of telephone sets would be astronomically expensive, to say the least and perhaps impossible. In response to this problem, the facilities needed to establish, and maintain a talking path between a pair of telephone sets provided in a common pool, to be used by a call when required and returned to the pool when no longer needed. This introduces the possibility that a system may be unable to set up a call on demand because of lack of available resources (e.g. telephone lines) at that time. Thus the question immediately arises: how much equipment must be provided so that the proportion of calls experiencing delays to be below a specified acceptable level?

Questions similar to that just posed arise in the design of many systems quite different from a telephone system. For example how many teletype writer stations can a time-shared computer serve?

All these questions share a common characteristic: in each case the times at which requests for service will occur and the lengths of time that these requests will occupy facilities cannot be predicted except in a statistical sense. The mathematical theory that studies the problem of design and analysis of such systems is known as Queueing Theory.

Every queue can be defined in terms of three characteristics: the input process, the service mechanism and the queue discipline. The input process describes the sequence of requests for service. For example, the input process is often specified in terms of the distribution of the lengths of time between consecutive customer arrival instants. For example, for a general distribution the letter G is used, M for the Exponential distribution (i.e. for a Poisson arrival process) and D for deterministic times. The service mechanism includes such characteristics as the number of servers and the lengths of time that the customers occupy the servers. The most common service times are, the exponentially distributed $\big(M\big)$ (i.e. for a Poisson arrival process, the deterministic times $\big(D\big)$, and the generally distributed times $\big(G\big)$.

The queue discipline specifies the way the blocked customers (customers who find all servers busy) are chosen later for service. A short hand notation to characterize a range of queueing models is the three-part code $A/B/c$. The first letter specifies the inter-arrival time distribution and the second one the service time distribution. The third term specifies the number of servers, while the notation can be extended with extra terms to cover more complicate queueing models.

1.2 Performance measures

Relevant performance measures in the analysis of a queueing model are:

- $Q_q(t)$: the number of waiting customers at time t.
- $Q_s(t)$: the number of customers in service at time t.
- $Q(t)$: the number of customers in the system (queue length) at time t.
- W_n: the waiting time of a customer.
- X_n: the service time of a customer.
- S_n: the sojourn time of a customer, which is equal to the waiting time plus the service time.

In particular, we are interested in mean performance measures, such as the mean waiting time and the mean sojourn time. Now, consider any queue with a state process that is regenerative, i.e, there are points that constitute a renewal process, where the state process of the system starts anew. Let the random variable $Q(t)$ denote the number of customers in the system at time t and let S_n denote the sojourn time of the n-th customer in the system. Under the assumption that the regenerative points of the system constitute a renewal process with (almost surely) finite aperiodic inter-renewal times, it can be shown that these random variables have limiting distributions as $t \to \infty$ and $n \to \infty$. These distributions are independent of the initial state of the system.

Let the random variables Q and S have the limiting distributions of $Q(t)$ and S_n respectively. We define

$$p_n = P(Q = n) = \lim_{t\to\infty} P(Q(t) = n),$$

$$F_s(x) = P(S \leq x) = \lim_{n\to\infty} P(S_n \leq x).$$

The probability p_n can be interpreted as the fraction of time that n customers are in the system and $F_s(x)$ gives the probability that the sojourn time of an arbitrary customer entering the system in not greater than x units of time. When $Q(t)$ is a regenerative process, which is the case in most applications, we have that:

$$\lim_{t\to\infty} \frac{1}{t}\int_0^t Q(x)dx = E(Q) \text{ and } \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} S_k = E(S) \text{with probability 1.}$$

So the long-run average number of customers in the system and the long-run average sojourn time are equal to $E(Q)$ and $E(S)$, respectively.

1.3 Utilization rate

If we denote by λ the arrival rate in a queueing system, and μ the maximum service capacity, then the utilization rate is essentially the ratio $\frac{\lambda}{\mu}$, i.e., it is the ratio of the rate at which 'work' enters the system over the maximum rate at which system can process this work. The work of an incoming customer is defined as the time of service he requires. In a single-server system $G/G/1$ with arrival rate λ and mean service time $\frac{1}{\mu}$, the amount of work arriving per unit time equals $\frac{\lambda}{\mu}$. To avoid that the queue length eventually grows to infinity, we have to require that $\frac{\lambda}{\mu} < 1$. It is common to use the notation $\rho = \frac{\lambda}{\mu}$. If $\rho < 1$, then ρ is called the occupation rate or server utilization, because it is the fraction of time the server is working.

1.4 Little's law

Little's law gives a very important relation between $E(Q), E(S)$ and the arrival rate λ. Little's law states that $E(Q) = \lambda E(S)$ or intuitively the long-run average number of customers in a stable system is equal to the long-run average effective arrival rate multiplied by the average time a customer spends in the system.

1.5 PASTA property

For queueing systems with Poisson arrivals, the very special property holds, that an arriving customer finds on average the same situation at the queueing system as an outside observer looking at the system at an arbitrary point in time. More precisely, the fraction of customers finding on arrival the system in some state A is exactly the same as the fraction of time the system is in the state A. This property is generally true only for Poisson arrivals. For this reason, it is known as Poisson Arrivals See Time Averages (PASTA) property.

1.6 The $M/M/1$ queue

In this chapter, we will analyze the model with a Poisson(λ) arrival process, exponential service times with mean $1/\mu$ and a single server. The queue discipline is First Come First Served (FCFS). We also require that the utilization rate $\rho = \lambda/\mu < 1$.

The exponential distribution allows for a very simple description of the state of the system at time t, namely the number of customers in the system. Neither do we have to remember when the last customer arrived, nor we have to register when the last customer entered service. Since, the exponential distribution is memoryless, this information does not yield a better prediction of the future. Let $X \sim Exp(\mu)$, then by the memoryless property, we get for all $x \geq 0$ and $t \geq 0$,

$$P(X > x + t | X > t) = P(X > x) = e^{-\mu x}.$$

We often use the memoryless property in the form

$$P(X < t + \Delta t | X > t) = 1 - e^{-\mu \Delta t} = \mu \Delta t + o(\Delta t), \ (\Delta t \to 0).$$

Based on the above property we get, for $\Delta t \to 0$,

$$\begin{aligned}
&p_0(t + \Delta t) = (1 - \lambda \Delta t) p_0(t) + \mu \Delta t p_1(t) + o(\Delta t), \\
&p_n(t + \Delta t) = \lambda \Delta t p_{n-1}(t) + (1 - (\lambda + \mu)\Delta t) p_n(t) + \mu \Delta t p_{n+1}(t) + o(\Delta t) \\
&n \geq 1.
\end{aligned}$$

Hence, by letting $\Delta t \to 0$, we obtain the following infinite set of differential equations for the probabilities $p_n(t)$.

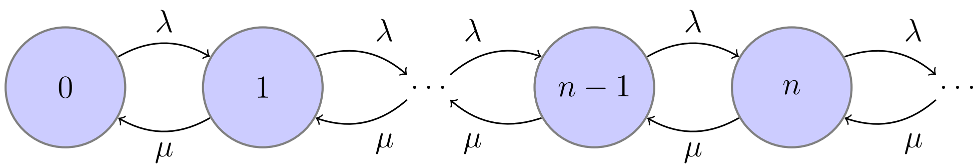
$$\begin{aligned}
p_0'(t) &= -\lambda p_0(t) + \mu p_1(t), \\
p_n'(t) &= \lambda p_{n-1}(t) - (\lambda + \mu) p_n(t) + \mu p_{n+1}(t), \qquad n \geq 1
\end{aligned} \tag{1.1}$$

Under the stability condition $\rho < 1$, we have that $p_n^{'}(t) \to 0$ and $p_n(t) \to p_n$, as $t \to \infty$. Hence from (1.1), it follows that the limiting or equilibrium probabilities p_n satisfy the equations:

$$\begin{aligned} 0 &= -\lambda p_0 + \mu p_1, \\ 0 &= \lambda p_{n-1} - (\lambda + \mu) p_n + \mu p_{n+1}, \qquad n \geq 1. \end{aligned} \tag{1.2}$$

In addition, the equilibrium probabilities p_n satisfy $\displaystyle\sum_{n=0}^{\infty} p_n = 1$

The transition rate diagram of the $M/M/1$ queue is:



The generic equation (1.2) is a second order recurrence relation with constant coefficients. Its general solution is of the form

$$p_n = c_1 x_1^n + c_2 x_2^n \quad , n \geq 0$$

where x_1 and x_2 are the roots of the quadratic equation

$$\lambda - (\lambda + \mu)x + \mu^2 = 0.$$

Solving the above equation yields $x = 1$ and

$$x = \frac{\lambda}{\mu} = \rho.$$

So, the generic solution is

$$p_n = c_1 + c_2 \rho^n, \; n \geq 0.$$

By the normalization equation

$$\sum_{n=0}^{\infty} p_n = 1$$

we conclude that c_1 must be 0, so again by the normalization equation we get that

$$c_2 = 1 - \rho.$$

So we conclude that

$$p_n = (1-\rho)\rho^n, \ \ n \geq 0$$

Alternatively, the limiting probabilities of the $M/M/1$ queue can be determined by applying the global balance principle, stating that for each set of states A, the flow out of set A is equal to the flow into that set. In fact, the equilibrium equations follow by applying this principle to a single state. So, if we apply the global balance principle to the set $A = \{0, 1, 2..., n-1\}$ we get the very simple relation

$$\lambda p_{n-1} = \mu p_n, n \geq 1.$$

Repeated application of this relation yields

$$p_n = \rho^n p_0, \ \ n \geq 1$$

so that, after the normalization, the solution

$$p_n = (1-\rho)\rho^n$$

, $n \geq 0$, follows.

From the equilibrium probabilities we can derive expressions for the mean number of customers in the system and the mean sojourn time in the system. For the first one we get

$$E(Q) = \sum_{n=0}^{\infty} nP(Q = n) = \sum_{n=0}^{\infty} np_n = \frac{\rho}{1-\rho}$$

and by applying the Little's law, we get that:

$$E(S) = \frac{E(Q)}{\lambda} = \frac{\rho}{\lambda\big(1-\rho\big)} = \frac{1}{\mu} \cdot \frac{1}{1-\rho}.$$

By the expressions of $E(Q)$ and $E(S)$ we observe that both quantities grow to infinity as $\rho \to 1$. This limiting behavior is caused by the variation in the arrival and service process. In fact, $E(Q)$ and $E(S)$ can also be determined directly by using Little's law and the PASTA property. Based on PASTA we know that the average number of customers in the system seen by an arriving customer equals $E(Q)$ and each of them has a service time with mean $1/\mu$. We mention that this is also true for the one in service, due to the memoryless property of the exponential distribution. So we get that the mean residual service time is $1/\mu$. Furthermore, a tagged customer has to wait for its own service time. Hence, we get the relation

$$E(S) = E(Q)\frac{1}{\mu} + \frac{1}{\mu}$$

and combing it with Little's law

$$E(Q) = \lambda E(S)$$

we find that

$$E(S) = \frac{1/\mu}{1-\rho}.$$

The mean number of customers in the queue, denoted by $E(Q_q)$ can be obtained from $E(Q)$ by subtracting the mean number of customers in service, so

$$E(Q_q) = E(Q) - \lambda/\mu = E(Q) - \rho = \frac{\rho}{1-\rho} - \rho = \frac{\rho^2}{1-\rho}.$$

The mean waiting time, $E(W)$, follows from $E(S)$ by subtracting the mean service time. This yields

$$E(W) = E(S) - \frac{1}{\mu} = \frac{\rho/\mu}{1-\rho}.$$

Denote by Q^a the number of customers in the system just before the arrival of a tagged customer and let B_k be the service time of the k-th customer. Also, the customer in service has a residual service time instead of an ordinary service time but these are identically distributed, since the exponential service time distribution is memoryless. So the r.vs B_k are independent and exponentially distributed with mean $1/\mu$. Then, we have that

$$S = \sum_{k=1}^{Q^a+1} B_k. \tag{1.3}$$

By conditioning on Q^a and using that Q^a and B_k are independent it follows that

$$P(S > t) = P\Big(\sum_{k=1}^{Q^a+1} B_k > t\Big) = \sum_{n=0}^{\infty} P\Big(\sum_{k=1}^{n+1} B_k > t\Big) P(Q^a = n). \tag{1.4}$$

So, we need to find the probability that an arriving customer finds n customers in the system. The PASTA property states that the fraction of customers finding on arrival n customers in the system is equal to the fraction of time there are n customers in the system, so $P(Q^a = n) = p_n = (1-\rho)\rho^n$. Using that

$$\sum_{k=1}^{n+1} B_k$$

is $Erlang -$ (n+1, μ) distributed, yields

$$\begin{aligned}
P(S>t) &= \sum_{n=0}^{\infty}\sum_{k=0}^{n}\frac{(\mu t)^k}{k!}e^{-\mu t}(1-\rho)\rho^n \\
&= \sum_{k=0}^{\infty}\sum_{n=k}^{\infty}\frac{(\mu t)^k}{k!}e^{-\mu t}(1-\rho)\rho^n \\
&= \sum_{k=0}^{\infty}\frac{(\mu\rho t)^k}{k!}e^{-\mu t} = e^{-\mu(1-\rho)t}, \quad t\ge 0.
\end{aligned}$$

Hence, S is exponentially distributed with parameter $\mu(1-\rho)$. This result can also be obtained via the use of transforms. Using that,

$$S = \sum_{k=1}^{Q^a+1} B_k,$$

and conditioning on Q^a yields:

$$\begin{aligned}
F_S^*(s) &= E(e^{-sS}) = \sum_{n=0}^{\infty} P(Q^a=n)E(e^{-s(B_1+\cdots+B_{n+1})}) \\
&= \sum_{n=0}^{\infty}(1-\rho)\rho^n E(e^{-sB_1})\cdots E(e^{-sB_{n+1}})
\end{aligned}$$

Since B_k is exponentially distributed with parameter μ, we have:

$$\begin{aligned}
E(e^{-sB_k}) &= \frac{\mu}{\mu+s}, \text{ so we get that} \\
F_S(s) &= \sum_{n=0}^{\infty}(1-\rho)\rho^n\Big(\frac{\mu}{\mu+s}\Big)^{n+1} = \frac{\mu(1-\rho)}{\mu(1-\rho)+s}
\end{aligned}$$

from which we can conclude that S is an exponential random variable with parameter $\mu(1-\rho)$. To find the distribution of the waiting time W, note that $S = W + B$, where the random variable B is the service time. Since W and B are independent, it follows that:

$$F_S(s) = F_W(s)\cdot F_B(s) = F_W(s)\cdot\frac{\mu}{\mu+s} \text{ and thus,}$$

$$F_W(s) = \frac{(1-\rho)(\mu+s)}{\mu(1-\rho)+s} = (1-\rho)+\rho\frac{\mu(1-\rho)}{\mu(1-\rho)+s}.$$

From the transform of W we conclude that W is equal to 0 with probability $(1-\rho)$, and with probability ρ is equal to an exponential random variable with parameter $\mu(1-\rho)$. Hence,

$$P(W>t) = \rho e^{-\mu(1-\rho)t},\ t\ge 0.$$

Note that

$$P(W > t|W > 0) = \frac{P(W > t)}{P(W > 0)} = e^{-\mu(1-\rho)t}$$

so the conditional waiting time $(W|W > 0)$ is exponentially distributed with parameter $\mu(1 - \rho)$.

1.7 Queueing Games

The economic analysis of queueing systems with strategic customers is an emerging tendency in the recent literature complementing the earlier studies in queueing theory, which concerned the performance evaluation of systems with passive (i.e. non-deciding) customers. In such a study, a certain reward cost structure is imposed on a queueing system that quantifies the customers' desire for service and their aversion to waiting. The customers are allowed to make decisions such as whether to join or balk, stay or renege, buy priority or not etc. and the questions we would typically like to answer are how do the customers behave in the system and what we can do to induce a desirable behavior. The collective behavior of the customers is analyzed as a game among the potential customers and we want to determine first the best response of a customer against a given strategy of the others and second the equilibrium strategies. Thus, Game theory plays a central role in this thesis. Therefore, we now overview the necessary background material.

In a strategic game, the players determine their action simultaneously and independently. In other words, a player chooses an action, he does not know the actions of the other players. Let $N = \{1, \ldots, n\}$ be a finite set of players and let A_i denote a set of actions available to player $i \in N$. A pure strategy for a player i is an action from A_i. A mixed strategy corresponds to a probability function which prescribes a randomized rule for selecting an action from A_i. We denote by S_i the set of strategies available to player i. A strategy profile $s = (s_1, s_2, \cdots, s_n)$ assigns a strategy $s_i \in S_i$ to each player $i \in N$. Each player is associated with a real payoff function $F_i(s)$. This function specifies the payoff received by player i, given that the strategy profile s is adopted by the players. We denote by s_{-i} a profile for the set of players $N \setminus \{i\}$. The function $F_i(s) = F_i(s_i, s{-}i)$ is assumed to be linear in s_i. This means that if s_i is a mixture with probabilities p and $1-p$ between strategies s_i^1 and s_i^2, then $F_i(s_i, s_{-i}) = pF_i(s_i^1, s_{-i}) + (1-p)F_i(s_i^2, s_{-i})$ for any s_{-i}.

Definition (1.1)**.** Strategy s_i^* is said to be a best response for player i against the profile s_{-i} if $s_i^* \in argmax_{s_i \in S_i} F_i(s_i, s_{-i})$

According to the definition mentioned above, we observe that in a non-cooperative game the player's payoff depends on the other customers' strategies. If the best response of a player does not depend on the other customers' strategies, then we have the following definition.

Definition (1.2)**.** Strategy s_i^1 is said to weakly dominate strategy s_i^2, if for any s_{-i}, $F_i(s_i^1, s_{-i}) \geq F_i(s_i^2, s_{-i})$ and for at least one s_i the inequality is strict. A strategys_i is weakly dominant if, regardless of what the other players do, the strategy ensures a player a payoff at least as high as any other strategy, and, the strategy earns a strictly higher payoff for some profile of other players' strategies. Hence, a strategy is weakly dominant if it is always at least as good as any other strategy, for any profile of other players' actions, and is strictly better for some profile of others' strategies. If a player has a weakly dominant strategy, then all other strategies are weakly dominated. If a strategy is always strictly better than all others for all profiles of other players' strategies, then it is strictly dominant.

In most cases there are no dominant strategies, so we are interested in finding a weaker form of solution, i.e., symmetric equilibrium strategies. Therefore, we give the following definition.

Definition (1.3)**.** A symmetric equilibrium strategy is a strategy $s^e \in S$ such that $\mathbf{s^e} \in argmax_{s \in S} F(s, \mathbf{s^e})$. In other words, s^e is a symmetric equilibrium if it is a best response against itself.

As aforementioned, the theory of strategic customers in queueing system deals with customers that make decisions in order to maximize their individual profits. The number of players is infinite and in the simplest case we suppose that they are homogeneous, i.e. all customers receive the same reward $\big(R\big)$ for having completed the service, they have the same cost $\big(C\big)$ per unit time in the system, and all customers have identical distributions of residual service times. Therefore, in considering the best response of a tagged customer, we assume that all the other customers follow the same strategy. According to this assumption, we are interested in finding only symmetric best responses. More specifically, in this framework we have the following definitions.

Definition (1.4)**.** A strategy of a customer is the set of all the rules that specify the decisions of the customer according to the circumstances of the game.

Definition (1.5)**.** A strategy a' is a best response of a tagged customer against a strategy a of the other customers in the system if $F(a', a) \geq F(b, a)$ for all available strategies b of the tagged-customer.

Definition (1.6)**.** A strategy a_1 weakly dominates strategy a_2 if $F(a_1, a) \geq F(a_2, a)$ for all available strategies a, and for at least one a the inequality is strict. A strategy a^* is said to be weakly dominant if it weakly dominates all the other strategies.

Definition (1.7)**.** A mixed strategy a_e is a symmetric equilibrium strategy if it is a best response against itself, i.e., $F(a_e, a_e) \geq F(a, a_e)$, for each strategy a.

In the following models we classify queues depending on whether or not their length can be observed. If an arriving customer in a system knows the number of customers in it, then the queue is called observable. In case that an arriving customer does not know the number of customers in the system, then the queue is called unobservable. Customers in both queueing models have to decide whether to join the system or not. So we consider the stochastic system as a non-cooperative game. Specifically, we assume the following:

(1) : When evaluating an individual's expected payoff, which is associated with a strategy $q^{'}$ as a response against all others using strategy q, we assume that steady-state conditions (based on all using strategy q) have been reached. In the following models there is an underlying Markov process, whose transition probabilities are induced by the common strategy selected by all customers. Hence, steady-state has the standard meaning of limiting probabilities and an individual assumes that this is the distribution over the states.

(2) : We compute the utility function of an incoming customer who decides to enter the system according to a strategy $q^{'}$ while all the others follow strategy q.

(3) : In case that a customer balks, then his utility function is zero.

(4) : We find the best response of the customer against the strategy q, by maximizing his utility function.

(5) : We find the equilibrium strategies of the customers if any.

1.8 Threshold Strategies

Suppose that an arriving customer has to choose between two actions, after observing the state of the system. A pure threshold with threshold n prescribes one of the actions, say A_1, for every state in $0, 1, \ldots, n-1$ and the other action, say A_2 otherwise.

Definition[1.8] A threshold strategy with threshold $x = n + p$, $n \in N$, $p \in [0, 1)$ prescribes mixing between the two pure threshold strategies n and $n + 1$ with weights $1 - p$ and p respectively. The resulting behavior is that all select a given action, say, A_1, when the state is $0 \leq i \leq n - 1$; select randomly between A_1 and A_2 when $i = n$, assigning probability p to A_1 and probability $1 - p$ to A_2; select A_2 when $i > n$. If x is an integer the strategy is pure, otherwise is mixed.

2 Strategic customer behavior in observable and unobservable queues

2.1 Naor's model - The observable case

In this chapter we will study the case in which the customers (players) observe the queue length and then they decide whether to join or not. We will also present various techniques for optimizing the customer's individual profit, the social welfare, and the system administrator's profit.

Historically, the strategic queueing theory began in 1964 when W.A Leeeman criticized the earlier studies in queueing theory. Specifically, Leeman claimed that the classical approach of queueing theory was intended for a planned economy and not for a capitalistic system as it should be. Therefore, he suggested the pricing of a queue for three main reasons:

- The optimization of the allocation of existing resources,
- The decentralization of administrative decisions,
- The planning of long-term investment decisions.

P. Naor added another main reason, which was the regulation of the demand. In 1969, Naor published the first paper in this area. The subject of the paper was the control of an FCFS $M|M|1$ queueing system. The assumptions of the Naor's paper are the following:

- The model is an $M/M/1$ queue with FCFS service discipline, with arrival and service rates λ and μ respectively.
- A customer earns a profit R from service completion, pays an entrance fee that is imposed by the system's manager and accumulates a cost C per unit of time staying in the system.

- Customers are risk neutral, that is they maximize the expected value of their utility.
- $R\mu \geq C$.
- The decision to join is irrevocable, i.e, reneging is not allowed.
- When a customer arrives he observes the queue length and then decides whether to join the queue or not.
- Information: Each customer at his instant arrival knows exactly the parameters and the number of the customers in the system.
- A customer who balks cannot return later.
- Utility functions of individual customers are identical and additive from the public (social) point of view.

2.1.1 Individual's optimization

According to the assumptions of the model, the stochastic system can be viewed as a game in which customers are considered as players. We assume that all customers follow a specific strategy $\vec{q} = (q_0, q_1, \ldots, q_n, \ldots)$, $q_n \in \{0, 1\}$, where, q_n is the joining probability when there are n customers in the system. Therefore, the system is a $M/M/1$ queue with service rate μ and variable arrival rate λq_n provided that there are n customers in the system, $n \geq 0$. If $\Pi_n(\vec{q})$, $n \in \mathbb{N}$ is the stationary distribution of the system, then we get that the real (effective) entrance rate in the system is $\lambda^* = \sum_{n=0}^{\infty} \lambda q_n \cdot \Pi_n(\vec{q})$. The mean waiting time of a customer depends on strategy $\vec{q}$ and according to Little's law and PASTA property is $\frac{E(Q)+1}{\mu}$. However, provided that a customer finds n customers in the system (an information that is available because we are in the observable case), then $E(Q) = n$. What is more, by applying the memoryless property and provided that queueing discipline is FCFS, the mean waiting time of the customer is $\frac{n+1}{\mu}$, a function that is independent of the strategy $\vec{q}$.

Let p be a specific admission fee. We consider that the customers in the system follow strategy $\vec{q}$. If an incoming customer follows $\vec{q'}$ strategy and observes n customers in the system, then independently of the other customers' strategy, his utility function will be

$$U(\vec{q'}, \vec{q}; n) = (1 - q_n') \cdot 0 + q_n'(R - p - C\frac{n+1}{\mu}).$$

We observe that $U(\vec{q'}, \vec{q}; n)$ is a linear function of q_n'. We also know that the best response of the n-tagged customer when the other customers follow $\vec{q}$ strategy is the solution of the maximizing problem $max_{q_n'} U(\vec{q'}, \vec{q}; n)$. Therefore we have the following cases:

$$BR(q) = \begin{cases} 0 & \text{when } R - p - C\frac{n+1}{\mu} < 0 \\ [0,1] & \text{when } R - p - C\frac{n+1}{\mu} = 0 \\ 1 & \text{when } R - p - C\frac{n+1}{\mu} > 0 \end{cases} \tag{1.5}$$

or, equivalently,

$$BR(q) = \begin{cases} 0 & \text{when } \frac{\mu(R-p)}{C} - 1 < n \\ [0,1] & \text{when } \frac{\mu(R-p)}{C} - 1 = n \\ 1 & \text{when } \frac{\mu(R-p)}{C} - 1 > n \end{cases} \tag{1.6}$$

Thus, a customer will join the queue if and only if, his net benefit is not negative. Equivalently, (1.6) proves that the customer who observes $n-1$ customers in front of him will join the system if

$$n < \frac{(R-p)\mu}{C}$$

Therefore, the highest number of customers that an incoming customer should observe in order to enter the system is $\left\lfloor \frac{(R-p)\mu}{C} \right\rfloor$. If we set $n_e = \left\lfloor \frac{(R-p)\mu}{C} \right\rfloor$, then the best response of an incoming customer is to join the queue if he finds at most $n_e - 1$ customers, and to balk if he finds more than $n_e - 1$ customers. We mention that in bibliography n_e is called Naor's threshold. More concretely, an equilibrium strategy has the form

$$\vec{q} = (1, 1, \ldots, 1[(n_e - 1) - \text{th position}], 0, 0, \ldots)$$

where 1 denotes that a customer joins the queue and 0 that a customer balks. Indeed, $\vec{q}$ is equilibrium strategy because independently of the other customers' strategies, a customer who follows the specific strategy has no incentive to change it. As a result, it can be concluded that the equilibrium strategy is a threshold strategy. Thus, the queue length could not be above the threshold n_e.

At this point we should mention that the role of the Nash-equilibrium strategy is not so important in Naor's model, because the decision of a customer that observes n customers in the system depends on his positive net benefit $\left(U(\vec{q'}, \vec{q}; n) = (1 - q_n') \cdot 0 + q_n'(R - p - C\frac{n+1}{\mu}\right)$ which is independent of the other customer's strategy $\vec{q}$. As a result, we get that the equilibrium strategy in the Naor's model is a dominant strategy.

2.1.2 Social benefit optimization

In this subsection, the goal is to find the optimal strategy that maximizes the mean social welfare. As social welfare we define the mean social reward per time unit minus the mean social cost per time unit. Intuitively, it is possible a customer's individual strategy to impose negative externalities on the total mean benefit of the customers in the system. Naor noticed in his paper 'The regulation of queue size by levying tolls' that the individual optimal threshold of a customer deviates from the socially optimal threshold.

An intuitive interpretation of this fact is that the decision of a customer to join induces longer delays for the future customers. On the other hand, these negative externalities are not taken into account in the case of individual optimization. Therefore, an individual may join the queue even if his own expected welfare is smaller than the expected reduction in welfare to future customers, which is not socially preferred.

Generally, in order to achieve social optimality we suppose that there is an administrative manager who accepts or forbids the entrance of customers in the system. Alternatively, the customers may be free to decide whether to enter or not, but the administrator may regulate the system using a pricing systems, i.e. by imposing an admission fee p that customers should pay in order to enter the system. We interpret this admission fee as server's profit. As the customers are aware of this entrance fee, then each one adopts the threshold strategy which maximizes his net benefit $\big($as it is described in the **individual profit maximization**$\big)$. Suppose that the administrator follows a specific strategy

$$\tilde{\mathbf{q}} = \big(q_0, q_1, \dots\big)$$

where q_n is the probability of accepting a customer, given that there are n customers in the system. Under this policy the system behaves as a $M/M/1$ queue with entrance rate $\lambda_n = \lambda q_n$, $n \geq 0$ and service rate μ. If $\Pi_n(\vec{q})$ denotes the steady state distribution of the system then

$$\Pi_n(\vec{q}) = B\rho^n q_o q_1 \cdots q_{n-1},$$

$n \in \mathbb{N}$, and $B = \Big(\sum_{n=0}^{\infty} \rho^n q_o q_1 \cdots q_{n-1}\Big)^{-1}$ with $\rho = \frac{\lambda}{\mu}$. If λ^* is the real entrance rate, then $\lambda^* = \sum_{n=0}^{\infty} \lambda q_n \Pi_n(\vec{q})$. We denote $E(Q)$ the mean number of customers in the system, thus, $E(Q) = \sum_{n=0}^{\infty} n\Pi_n(\vec{q})$. We notice that the additional payments need not be considered part of the social welfare, because transactions between the members of the society do not influence the social welfare. Therefore, the mean social welfare per time unit is

$$S(\vec{q}) = \lambda^*(R - p) - CE(Q) + \lambda^* p = \lambda^* R - CE(Q)$$

We want to determine the optimal strategy $\vec{q}$ in order to maximize the social welfare. So, we need to solve the problem

$$S_{soc}^{optimal} = \max_{\vec{q}} S(\vec{q}).$$

Because we are in the observable case, the decision framework is that of Dynamic programming (Markov decision processes) with perfect information. Therefore, the optimal policies are deterministic. So, we will limit our search of optimal policies within the class of pure policies. Specifically, under the assumption that $n \geq 0$, is the number of customers in the system, then $q_n = 1$ if he accepts the entrance of an incoming customer whereas $q_n = 0$ if he denies it. Let $\vec{q} = (q_0, q_1, q_2, \dots)$, $q_i \in \{0, 1\}$ be a specific pure strategy of the administrator. Thus, we suppose that he will observe the number of customers in the system and will accept the entrance only for the customers that induce positive welfare. As the number of customers in the system increases, then the total waiting time of all customers increases too. So the positive effect of the entrance of a tagged-customer on the social welfare decreases. The first time that an entrance of an incoming customer implies a net decrease on the overall social welfare the administrator will deny his entrance, because of the convexity of the mean waiting time at the arrival rate. We observe that if we find that n, which determines the turning point from $q_{n-1} = 1$ to $q_n = 0$, then simultaneously the strategy of the administrator is precisely determined. Specifically, we set

$$n^* = n^*\big(\vec{q}\big) = \min\big\{n : q_n = 0\big\}.$$

We have that $n^*(\vec{q}) = \infty$ in case that $q_n = 1 \ \ \forall n \in \mathbb{N}$, so the system behaves as a $M/M/1$ queue, and $n^*(\vec{q}) < \infty$ in case that exists at least one n such that $q_n = 0$ so the system behaves as a $M/M/1/n^*$.

However, under a specific value of p and for any strategy $\vec{q}$ the socially optimum threshold n^* should satisfy

$$n^* \leq \frac{(R-p)\mu}{C}.$$

Indeed, if $n^* > \frac{(R-p)\mu}{C} \iff R - p - \frac{n^* C}{\mu} < 0$, then there are customers (at least the last one) who enter the system and brings about negative welfare which is not socially preferred. Thus, under a specific value of p the maximum possible socially optimal threshold is smaller than the individual optimal Naor threshold $\left\lfloor \frac{\mu(R-p)}{C} \right\rfloor$.

Then, the queueing system behaves as a $M/M/1/n^*$ model. The stationary distribution under the assumption that $\rho \neq 1$, is

$$p_k = \frac{(1-\rho)\rho^k}{1-\rho^{n^*+1}},\ \ 0 \le k \le n^*.$$

Thus, the real entrance rate is

$$\lambda^* = \sum_{k=0}^{n^*-1} \lambda p_k = \lambda(1-p_{n^*}) = \lambda \frac{1-\rho^{n^*}}{1-\rho^{n^*+1}}$$

and the expected queue length is

$$E(Q) = \sum_{k=0}^{n^*} k p_k = \frac{\rho}{1-\rho} - \frac{(n^*+1)\rho^{n^*+1}}{1-\rho^{n^*+1}}.$$

Let $S(n^*)$ be the social benefit per time unit if n^* is the entrance threshold. Then,

$$S(n^*) = R\lambda^* - CE(Q) = R\lambda \frac{1-\rho^{n^*}}{1-\rho^{n^*+1}} - C\Big[\frac{\rho}{1-\rho} - \frac{(n^*+1)\rho^{n^*+1}}{1-\rho^{n^*+1}}\Big]. \quad (1.7)$$

Our objective is to find the socially optimal social threshold n^*. Based on (1.7) we compute the difference

$$S(n^*) - S(n^*-1).$$

Under the assumption that $\rho < 1$, and by doing some algebra we get that

$$S(n^*) - S(n^*-1) = \frac{\lambda R(1-\rho)^2\rho^{n^*-1}}{(1-\rho^{n^*+1})(1-\rho^{n^*})} + \frac{C\Big((n^*+1)\rho - \rho^{n^*+1} - n^*\Big)\rho^{n^*}}{(1-\rho^{n^*+1})(1-\rho^{n^*})}$$

We will prove that $S(n)$ is a concave function of $n \in \mathbb{N}$.

$$\begin{aligned}
S(n^*) - S(n^*-1) &= \lambda R \frac{1-\rho^{n^*}}{1-\rho^{n^*+1}} - \lambda R \frac{1-\rho^{n^*+1}}{1-\rho^{n^*}} + C\frac{(n^*+1)\rho^{n^*+1}}{1-\rho^{n^*+1}} - C\frac{n\rho^{n^*}}{1-\rho^{n^*}} \\
&= \lambda R \frac{1-2\rho^{n^*}+\rho^{2n^*}-1+\rho^{n^*-1}+\rho^{n^*+1}-\rho^{2n^*}}{(1-\rho^{n^*+1})(1-\rho^{n^*})} \\
&+ C\frac{(n^*+1)\rho^{n^*+1}-(n^*+1)\rho^{2n^*+1}-n\rho^{n^*}-n^*\rho^{2n^*+1}}{(1-\rho^{n^*+1})(1-\rho^{n^*})} \\
&= \frac{\lambda R(1-\rho)^2\rho^{n^*-1}}{(1-\rho^{n^*+1})(1-\rho^{n^*})} + \frac{C\Big((n^*+1)\rho - \rho^{n^*+1} - n^*\Big)p^n}{(1-\rho^{n^*+1})(1-\rho^{n^*})}. \quad (1.8)
\end{aligned}$$

We have,

$$\begin{aligned}
S(n^*) - S(n^*-1) &\geq 0 \\
&\Leftrightarrow \quad \lambda R(1-\rho)^2 \geq C\rho(n^* + \rho^{n^*+1} - (n^*+1)) \\
&\Leftrightarrow^{\rho=\lambda/\mu} \quad \mu R(1-\rho)^2 \geq C(n^* + \rho^{n^*+1} - (n^*+1)\rho) \\
&\Leftrightarrow \quad \frac{R\mu}{C} \geq \frac{n^* + \rho^{n^*+1} - (n^*+1)\rho}{(1-\rho^2)} \qquad (1.9)
\end{aligned}$$

We consider the function

$$g(n^*) = \frac{n^* + \rho^{n^*+1} - (n^*+1)\rho}{(1-\rho)^2}$$

By using algebra we get that

$$g(n^*) = \frac{1}{1-\rho}\sum_{k=1}^{n^*}(1-\rho^k).$$

Obviously, $g(n^*)$ is an increasing function of n^*. So there exist a point $n_0^* \in \mathbb{N}$ such that

$$g(n^*) \leq \frac{R\mu}{C} \quad \forall n^* \leq n_0^*$$

and

$$g(n^*) > \frac{R\mu}{C} \quad \forall\, n^* > n_0^*.$$

We determine the social threshold to be that point, $n_{soc} = n_0^*$. Therefore, for

$$n^* \leq n_{soc}, \quad S(n^*) - S(n^*-1) \geq 0$$

and for

$$n^* > n_{soc}, \quad S(n^*) - S(n^*-1) \leq 0.$$

By the previous property, we find that

$$n_{soc} = \max\left\{n^* : g(n^*) \leq \frac{R\mu}{C}\right\}, \qquad (1.10)$$

is the unique maximum of $S(n^*)$.

Thus, as the n_{soc} is known and given by (1.10) the system's administrator can derive it by imposing an appropriate admission fee p_{soc} so that $\left\lfloor \frac{\mu(R-p_{soc})}{C} \right\rfloor = n_{soc}$ or equivalently the social optimization can be achieved

$$\forall p_{soc} \in \left[R - \frac{C(n_{soc}-1)}{\mu}, R - \frac{Cn_{soc}}{\mu}\right].$$

Lemma 2.1: The social optimal threshold n_{soc} is smaller than the individual optimal threshold.

Proof. We consider the function

$$h : \mathbb{N} \to \mathbb{R}^{+} \text{with } \ h(n) = g(n) - n.$$

The function h is positive. Indeed,

$$\begin{aligned} h(n) &\geq 0 \\ &\Leftrightarrow g(n) - n \geq 0 \\ &\Leftrightarrow \frac{1}{1-p}\sum_{k=1}^{n}(1-p^{k}) - \frac{1}{1-p}n(1-p) \geq 0 \\ &\Leftrightarrow \frac{1}{1-p}\sum_{k=1}^{n}(1-p^{k}-1+p) \geq 0 \\ &\Leftrightarrow \frac{p}{1-p}\sum_{k=1}^{n}(1-p^{k-1}) \geq 0 \ \text{ which holds} \forall n \in \mathbf{N}. \end{aligned}$$

Therefore, we have that $g(n_{soc}) \geq n_{soc}$. By (1.10) we obtain that

$$g(n_{soc}) \leq \frac{R\mu}{C} \ \text{ and } \ n_{soc} \leq \frac{R\mu}{C} \Leftrightarrow n_{soc} \leq n_{e}$$

$\square$

Remark We proved that for the socially optimal threshold n^{*} and the corresponding price $p(n^{*})$, we have that $R - p(n^{*}) - \frac{C(n+1)}{\mu} > 0$ for $n \leq n^{*}$. Therefore, all the incoming customers have a positive welfare, thus there is not a unique p so that the administrative manager earns all the customer's profit. The reason is that each customer knows exactly the number of customers in the system so his net benefit differs from the other customers'. We will see in the Unobservable case that an administrative manager can achieve to earn all the positive welfare of the customers by imposing a single admission fee p.

Remark Instead of imposing an admission fee in order to motivate customers to adopt the social threshold, we could regulate the queue by imposing a toll on waiting, i.e by increasing the cost per time unit in the system by a positive number $t \in \mathbb{R}^{+}$. Such a toll t, induces the optimal threshold n^{*} if, $n^{*} = \left\lfloor \frac{R\mu}{C+t} \right\rfloor, \ t \geq 0$. The social optimality can be also achieved without imposing any admission fees, or additional tolls. Specifically, R.Hassin suggested in his paper "Consumer information in markets with random products quality: The case of queues and balking," that LCFS-PR leads to a socially optimal behavior by the customers.

2.1.3 Manager's profit optimization

In this section we will study the Naor's problem under the view of the administrative manager. Specifically, we would like to find a way to maximize the manager's profit per time unit.

We assume that an entrance fee p is imposed, which customers should pay in order to enter the system. This fee is is considered to be manager's profit. Since the customers know the value of p, then as in the case of 'Individual's profit maximization' case, they follow the threshold strategy

$$n_e(p) = \Big\lfloor \frac{(R-p)\mu}{C} \Big\rfloor. \tag{1.11}$$

Thus the system will be a $M/M/1/n_e(p)$ queue with entrance rate λ and service rate μ. The stationary distribution under the assumption that $\rho \neq 1$, is

$$p_k = \frac{(1-\rho)\rho^k}{1-\rho^{n^e(p)+1}}, \ 0 \leq k \leq n^e(p)$$

and the real entrance rate is

$$\lambda^* = \sum_{k=0}^{n^e(p)-1} \lambda p_k = \lambda(1 - p_{n^e(p)}) = \lambda \frac{1-\rho^{n^e(p)}}{1-\rho^{n^e(p)+1}}.$$

The manager's profit depends only in the real entrance rate, thus under a specific strategy n_e we have that the manager's profit per time unit is:

$$Z = \lambda \frac{1-\rho^{n^e(p)}}{1-\rho^{n^e(p)+1}} p \tag{1.12}$$

The goal is to find the p, which not only will give incentive to the customers to enter the queue, but also will be the highest possible one (under the interpretation of manager's profit). Under a specific individual threshold n_e, according to (1.11) we get that

$$R - \frac{(n_e+1)C}{\mu} < p \leq R - \frac{n_e C}{\mu}. \tag{1.13}$$

As the administrative manager wants to impose the maximum possible price under strategy n_e, then he sets

$$p(n_e) = R - \frac{n_e C}{\mu}. \tag{1.14}$$

Therefore, instead of finding the optimal price p, we can find threshold n_m which maximizes the manager's profit., thus (1.14) yields that $p = R - \frac{nC}{\mu}$.

From (1.12) and (1.14) under a specific threshold n we have that the manager's profit per time unit is

$$\begin{aligned} Z(n) &= \lambda \cdot \frac{1-\rho^n}{1-\rho^{n+1}}\Big(R - \frac{Cn}{\mu}\Big) \\ &= \lambda \cdot \frac{1-\rho^n}{1-\rho^{n+1}} R\Big(1 - \frac{Cn}{R\cdot\mu}\Big) \\ &= \lambda \cdot R \frac{1-\rho^n}{1-\rho^{n+1}}\Big(\frac{V_e - n}{V_e}\Big) \text{ where } V_e = \frac{R\mu}{C} \end{aligned} \tag{1.15}$$

Theorem (2.1)**.** *The manager's threshold n_m, is derived by the integer part of the solution of the equation $g(x) = 0$ where $g : [0, +\infty) \Rightarrow \mathbb{R}$ and*

$$g(x) = x + \frac{(1-\rho^{x+1})^2}{\rho^x(1-\rho)^2} - \frac{R\mu}{C},\ \rho \in (0,1)$$

.

Proof.

In order to find n_m we must solve the maximizing problem $\max_n Z(n)$. If $Z\big(n\big) \geq Z\big(n-1\big)$ then we have that:

$$\begin{aligned} \lambda R \frac{1-\rho^n}{1-\rho^{n+1}}\Big[\frac{V_e - n}{V_e}\Big] &\geq \lambda R \frac{1-\rho^{n-1}}{1-\rho^n}\Big[\frac{V_e - (n-1)}{V_e}\Big] \iff \\ \big(V_e - n + 1\big)\Big(\frac{1-\rho^n}{1-\rho^{n+1}} - \frac{1-\rho^{n-1}}{1-\rho^n}\Big) &\geq \frac{1-\rho^n}{1-\rho^{n+1}} \iff \\ \big(V_e - n + 1\big)\frac{\big(1-\rho^n\big)^2 - \big(1-\rho^{n-1}\big)\big(1-\rho^{n+1}\big)}{\big(1-\rho^n\big)\big(1-\rho^{n+1}\big)} &\geq \frac{1-\rho^n}{1-\rho^{n+1}} \iff \\ \big(V_e - n + 1\big)\frac{\rho^{n-1}\big(1-\rho\big)^2}{\big(1-\rho^n\big)\big(1-\rho^{n+1}\big)} &\geq \frac{1-\rho^n}{1-\rho^{n+1}} \iff \\ \big(V_e - n + 1\big) &\geq \frac{\big(1-\rho^n\big)^2}{\rho^{n-1}\big(1-\rho\big)^2}. \end{aligned} \tag{1.16}$$

If $Z(n) > Z(n+1)$ then

$$
\begin{aligned}
\lambda R\frac{1-\rho^n}{1-\rho^{n+1}}[\frac{V_e-n}{V_e}] &> \lambda R\frac{1-\rho^{n+1}}{1-\rho^{n+2}}[\frac{V_e-\big(n+1\big)}{V_e}] \iff \\
\big(V_e-n\big)\left(\frac{1-\rho^n}{1-\rho^{n+1}}-\frac{1-\rho^{n+1}}{1-\rho^{n+2}}\right) &> -\frac{1-\rho^{n+1}}{1-\rho^{n+2}} \iff \\
\big(V_e-n\big)\frac{-\rho^n\big(1-\rho\big)^2}{\big(1-\rho^{n+1}\big)\big(1-\rho^{n+2}\big)} &> -\frac{1-\rho^{n+1}}{1-\rho^{n+2}} \iff \\
\big(V_e-n\big)\frac{\rho^n\big(1-\rho\big)^2}{\big(1-\rho^{n+1}\big)\big(1-\rho^{n+2}\big)} &< \frac{1-\rho^{n+1}}{1-\rho^{n+2}} \iff (0<\rho<1) \\
\big(V_e-n\big) &< \frac{\big(1-\rho^{n+1}\big)^2}{\rho^n\big(1-\rho\big)^2}. \qquad (1.17)
\end{aligned}
$$

As $V_e=\frac{R\mu}{C}$ then we get that

$$
n-1+\frac{\big(1-\rho^n\big)^2}{\rho^{n-1}\big(1-\rho\big)^2}\leq\frac{R\mu}{C}<n+\frac{\big(1-\rho^{n+1}\big)^2}{\rho^n\big(1-\rho\big)^2} \quad . \qquad (1.18)
$$

We consider the function

$$
g\big(x\big)=x-1+\frac{\big(1-\rho^x\big)^2}{\rho^{x-1}\big(1-\rho\big)^2}, \quad x\in[0,\infty), \quad 0<\rho<1.
$$

Then:

$\big(i\big)$: $g\big(0\big)=-1$

$\big(ii\big)$: $g\in C^\infty$

$\big(iii\big)$: $g\big(x\big)=x-1+\frac{\rho^{x+1}-2\rho+\rho^{-x-1}}{\big(1-\rho\big)^2}$

$\big(iv\big)$: $g'\big(x\big)=1+\frac{1}{\big(1-\rho\big)^2}\ln\rho\big(\rho^{x+1}-\rho^{-x+1}\big)=1+\frac{1}{\big(1-\rho\big)^2}\ln\rho\frac{\rho^{2x}-1}{\rho^{x-1}}>0$

as $0<\rho<1$, we get that $\ln\rho<0$, $\rho^{2x}<1$. Afterwards, we consider the function

$$
f\big(x\big)=g\big(x\big)-\frac{R\mu}{C}.
$$

We observe that

- f is continuous in $[0,\infty)$,

- $f'\big(x\big) > 0 \ \ \forall x \in R_0^+$ since $g^{'}(x) > 0$ and
  $f\big(0\big) = g\big(0\big) - \frac{R\mu}{C} = -1 - \frac{R\mu}{C} \leq 0,$

- $\lim_{x \to \infty} f(x) = \infty,$ and $\lim_{x \to \infty} \rho^x = 0$. Therefore, there is a unique root, such that

$$f\big(x_0\big) = 0 \iff g\big(x_0\big) = \frac{R\mu}{C}.$$

  By denoting the manager's threshold

$$n_m = \lfloor x_0 \rfloor$$

  we get that $n_m \leq x_0 < n_{m+1}$ thus as g is increasing in n this immediately yields

$$g\big(n_0\big) \leq g\big(x_0\big) = \frac{R\mu}{C} < g\big(n_{m+1}\big)$$

  so, the threshold n_m satisfy the double inequality $\big(1.18\big)$.

$\square$

Theorem (2.2)**.** *If n_e is the individual threshold, n_{soc} is the social threshold and n_m the threshold in manager's profit, then*

$$n_m \leq n_{soc} \leq n_e.$$

2.2 Edelson's and Hildebrand's model - Unobservable case

The properties of the basic unobservable single server queue were discovered by Edelson and Hildebrand.The assumptions of the model are:

$\big(1\big)$: Customers arrive according to a Poisson$\big(\lambda\big)$ distribution.

$\big(2\big)$: The service times are exponentially distributed.

$\big(3\big)$: There is one server in the system.

$\big(4\big)$: The service discipline is FCFS.

(5) : There is a reward R that each customer earns from his service completion.

(6) : There is a cost C per unit time a customer stays in the system.

(7) : There is an entrance fee p that is imposed by the administrative manager which the customers have to pay in order to enter the system $\big(0 \leq p < R\big)$.

(8) : Each customer decides whether to join the system or balk, with probability $q \in [0, 1]$

(9) : If a customer balks, then his utility is zero.

2.2.1 Individual's optimization

In the unobservable case, we suppose that there is an admission fee p, of which the customers are aware. Under this information, all customers adopt the same mixed strategy that can be described by a number q, $0 \leq q \leq 1$, which is the probability of joining the system. Thus, the system behaves as an $M/M/1$ queue with service rate μ and arrival rate λq. Also, under the stability condition $\lambda q < \mu$, the mean waiting time of a customer in the system is $\frac{C}{\mu - \lambda q}$. Thus, the net benefit of a tagged customer who joins the queue with probability $q^{'}$, while the others follow q strategy is,

$$U\big(q^{'}, q\big) = (1 - q^{'}) \cdot 0 + q^{'}\Big(R - p - \frac{C}{\mu - \lambda q}\Big).$$

In order to determine the best response of a customer we need to solve the maximizing problem $\max_{q^{'}} U(q^{'}, q)$ under the restriction $q \in [0, q_{max}]$, with $q_{max} = \min(1, \frac{\mu}{\lambda})$. We observe that the utility function of a customer is a linear function of $q^{'}$ therefore we have the following cases:

$$BR(q) = \begin{cases} 0 & \text{when } R - p - C\frac{1}{\mu - \lambda q} < 0 \\ [0, q_{max}] & \text{when } R - p - C\frac{1}{\mu - \lambda q} = 0 \\ q_{max} & \text{when } R - p - C\frac{1}{\mu - \lambda q} > 0. \end{cases} \tag{1.19}$$

Or equivalently,

$$BR(q) = \begin{cases} 0 & \text{when } q > \frac{1}{\lambda}(\mu - \frac{C}{R-p}) \\ [0, q_{max}] & \text{when } q = \frac{1}{\lambda}(\mu - \frac{C}{R-p}) \\ q_{max} & \text{when } q < \frac{1}{\lambda}(\mu - \frac{C}{R-p}). \end{cases} \tag{1.20}$$

Theorem (2.3)**.** *In the unobservable $M/M/1$ queue with FCFS discipline, an equilibrium strategy always exists and it is unique. Then, we have the following cases:*

Case I: If $R - p - \frac{C}{\mu} \leq 0$, then the $q_e = 0$ is the only equilibrium strategy.

Case II: If $R-p-\frac{C}{\mu-\lambda q_{max}} \geq 0$, then $q_e = q_{max}$ is the only equilibrium strategy.

Case III: If $R-p-\frac{C}{\mu-\lambda\cdot 0} < 0 < R-p-\frac{C}{\mu-\lambda q_{max}}$ then, there is a unique equilibrium strategy $q_e = \frac{1}{\lambda}(\mu-\frac{C}{R-p})$, which is the solution of the equation $R-p-\frac{C}{\mu-\lambda q} = 0$.

Proof. :

Case I: If $R-p-\frac{C}{\mu} < 0$, i.e. even if the system is empty, a customer that decides to enter the system, has a negative mean net benefit. Therefore, his best response is to balk ($q^{'} = 0$)which is a best response against itself and so, an equilibrium strategy. In case that $R-p-\frac{c}{\mu} = 0$ then the customer is indifferent about joining the system or balking, thus every $q^{'} \in [0, q_{max}]$ is best response. But only $q^{'} = 0$ is the best response against itself and thereafter $q_e = 0$ is the only equilibrium strategy.

Case III : In case that $R-p-\frac{C}{\mu-\lambda q_{max}} \geq 0$, or equivalently, even if all potential customers join the queue, a customer who decides to join the system has a non-negative benefit. Therefore, the best response of a customer is $q^{'} = q_{max}$, thus $q_e = q_{max}$ is the unique equilibrium strategy.
Case II :

$$R-p-\frac{C}{\mu-\lambda q_{max}} < 0 < R-p-\frac{C}{\mu} \quad (*).$$

As $R-p-\frac{C}{\mu-\lambda q_{max}} < 0$ then $q^{'} = q_{max}$ is not the best response. Moreover, as $R-p-\frac{C}{\mu} > 0$ then a customer who balks will loose a positive net benefit, therefore $q^{'} = 0$ is not the best response.
We know that

$$W(q) = \frac{C}{\mu-\lambda q}$$

is an increasing function of q, therefore, $R-p-\frac{C}{\mu-\lambda q}$ is decreasing in q. Also, $R-p-\frac{C}{\mu-\lambda q}$ is a continuous function of q. Then by $(*)$, exists a unique root q_e, such that $R-p = \frac{C}{\mu-\lambda q} \iff q_e = \frac{1}{\lambda}(\mu-\frac{C}{R-p})$. As $R-p-\frac{C}{\mu-\lambda q}$ is a decreasing function of q, then $\forall q \in [0, q_e]$ is $R-p-\frac{C}{\mu-\lambda q} > 0$, and the best response of an incoming customer is $q^{'} = q_{max}$. For each $q > q_e$ we have that

$R-p-\frac{C}{\mu-\lambda q}<0$, therefore, the best response for an incoming customer is $q^{'}=0$. As $R-p-\frac{C}{\mu-\lambda q_e}=0$ then the incoming customer is indifferent about joining the system or balking thus his best response is $q^{'}\in[0,q_{max}]$. However, the unique equilibrium strategy is $q^{'}=q_e$. □

2.2.2 Social Optimization

As a customer who joins the queue imposes negative externalities on the others, individual's optimization leads to excessive congestion. We suppose that there is an administrative manager (e.g. the government) that wants to find a way in order to socially regulate the queueing system without observing the system. We consider that the administrative manager accepts the entrance of the customers with probability q (it is the same for all the customers because we are in the unobservable case). We denote by $S\big(q\big)$ the sum of customer's welfare and the welfare of the service system. Therefore,

$$S\big(q\big)=\lambda q\big(R-p-\frac{C}{\mu-\lambda q}\big)+\lambda qp=\lambda q\big(R-\frac{C}{\mu-\lambda q}\big).$$

The goal is to find the socially optimal joining rate λq_{soc} in order to maximize the social welfare. Under the stable condition $\lambda q<\mu$, in order to find q_{soc} we need to solve the maximizing problem

$$\max_{0\le q\le 1} S\big(q\big)=S\big(q_{soc}\big),$$

$$q\in[0,q_{max}=min\big\{1,\frac{\mu}{\lambda}\big\}].$$

We have that,

$$\frac{\partial S\big(q\big)}{\partial q} \;=\; \lambda\left(R-\frac{C}{\mu-\lambda q}\right)-\lambda^2 q\frac{c}{\big(\mu-\lambda q\big)^2}$$

and,

$$\frac{\partial^2 S\big(q\big)}{\partial q^2} \;=\; -\frac{2c\mu\lambda^2}{\big(\mu-\lambda q\big)^3}. \tag{1.21}$$

Therefore, we observe that $S\big(q\big)$ is a concave function of q, $q\in[0,q_{max}]$ so, every local maximum of S is a global maximum. We will prove that,

$$q_{soc}=\begin{cases}0 & \text{when } S'\big(0\big)\le 0\\ q_{max} & \text{when } S'\big(q_{max}\big)\ge 0\\ \frac{1}{\lambda}\big(\mu-\sqrt{\frac{C\mu}{R}}\big) & \text{when } S'\big(q_{max}\big)<0<S'\big(0\big)\end{cases} \tag{1.22}$$

Indeed, as $S\big(q\big)$ is concave and differentiable in q its derivative is decreasing in q. Thus,

- If $S'\big(0\big) \leq 0$ then $S'\big(q\big) \leq 0 \quad \forall q \in [0,1]$, so $S\big(q\big)$ is a decreasing function in q, therefore $S(0)$ is maximum. Thus, the optimal social strategy is $q_{soc} = 0$
- If $S'\big(q_{max}\big) \geq 0$ then $S'\big(q\big) \geq 0$ so $S\big(q\big)$ is increasing in q, therefore for $q = q_{max}$, $S(q)$ is maximized. Thus, $q_{soc} = q_{max}$.
- If $S'\big(q_{max}\big) < 0 < S'\big(0\big)$ then by the continuity and monotonicity of S in q, there exists a unique root q^* such that $S'\big(q^*\big) = 0$. Thus, $S(q^*)$ is a global maximum of S. Therefore, $q_{soc} = q^* = \frac{1}{\lambda}\big(\mu - \sqrt{\frac{C\mu}{R}}\big)$.

  We observe that that if $p = 0$ we get that $q_e(0) \geq q_{soc}$, i.e., in the unobservable case the number of customers in the system under social optimization is always smaller than the number of customers in the system under individual maximization when there is not any admission fee.

  A question that may arise here is to determine the value of p that $q_{soc} = q^e(p)$ holds. This idea will be analyzed in the case of profit maximization.

2.2.3 Profit maximization

In this case we consider a monopolistic server that sets an admission fee p, the goal of which is to maximize his profit. Having evaluating his utility function, a customer determines his best response and his equilibrium strategies. Thus, as in the case of individual optimization, we get that the equilibrium strategy is

$$q_e(p) = \begin{cases} 0 & \text{when } R - p \leq \frac{C}{\mu} \\ \frac{1}{\lambda}\big(\mu - \frac{C}{R-p}\big) & \text{when } \frac{C}{\mu} < R - p < \frac{C}{\mu-\lambda} \\ q_{max} & \text{when } R - p \geq \frac{C}{\mu-\lambda} \end{cases} \tag{1.23}$$

Equivalently, if we solve the above inequalities in p, we get that

$$q_e(p) = \begin{cases} 0 & \text{when } p \geq R - \frac{C}{\mu} \\ \frac{1}{\lambda}\big(\mu - \frac{C}{R-p}\big) & \text{when } R - \frac{C}{\mu-\lambda} < p < R - \frac{C}{\mu} \\ q_{max} & \text{when } p \leq R - \frac{C}{\mu-\lambda} \end{cases} \tag{1.24}$$

The idea is that in order to find the appropriate p which maximizes the manager's profit, the manager can urge the customers to adopt that strategy which achieves his profit optimization, as it can be shown by (1.24) and (1.25). Thus, manager's profit per time unit is $S_m = p \cdot \lambda q_e(p)$.

Note that the equilibrium probability $q_e(p)$ is a decreasing function of p. If we set $p_1 = R - \frac{C}{\mu}$ and $p_2 = R - \frac{C}{\mu-\lambda}$ then we have the following cases:

- If $p < p_2$ then $q_e(p) = q_{max}$ and $S_m = \lambda q_{max} p < \lambda q_{max} p_1$ therefore, the values of $p \in (0, p_1)$ are not optimal.

- If $p > p_1$ then $q_e(p) = 0$, so $S_m = 0$.

Therefore, we consider that $p \in [p_1, p_2]$ or equivalently $\frac{c}{\mu} < R - p < \frac{c}{\mu-\lambda}$.When $p \in [p_1, p_2]$, we proved that the individual's equilibrium strategy is the unique solution of the equation $U(q_e(p)) = 0$. Therefore, $q_e(p) = \frac{1}{\lambda}\big(\mu - \frac{C}{R-p}\big) \iff$

$$p_e(q) = R - C\frac{1}{\mu - \lambda q}.$$

We note that under equilibrium strategy, if we set $p = p_e(q)$ then the percentage of customers that enter the system is q. As a result, we get that

$$S_m = \lambda p_e(q) q = \lambda q (R - C\frac{1}{\mu - \lambda q}).$$

Based on the last equation we mention some observations:

- We can see that manager's utility function coincides with the social welfare. Thus, we can derive the optimal q with the same procedure as in the social optimization case.

- The reasons why the manager's profit and the social welfare coincides are the homogeneity of the customers and the fact that they cannot observe the queue length.

- The manager leaves zero surplus to the customers, thus, the whole social welfare becomes his personal profit.

- The optimal admission fee $p_m = R - \frac{C}{\mu - \lambda q_{soc}}$, is a decreasing function of λ. This means that an increase in the demand induces an increase in the customers' waiting time. Due to the increase in their waiting time, customers perceive their service as a lower quality product, something that pushes the manager to decrease the admission fee.

3 Bridging the observable and the unobservable cases: Overview

3.1 Introduction

Hassin in his paper "Consumer information in markets with random products quality: The case of queues and balking" compares social welfare and profit maximization in the observable and unobservable models. Let S^u_{soc}, S^o_{soc} denote the social welfare under a social welfare-maximizing policy in the unobservable and observable models, respectively. Similarly, S^u_m, S^o_m denote the profit under a profit-maximizing admission fee. We have proved that

$$S^u_{soc} = S^u_m.$$

Let $S^{o'}_{soc}$ denote the social welfare in the observable model when a profit maximizing admission fee is charged. We have mentioned that profit-maximizing fee in the unobservable case is the same as the optimal social fee. In case that the queue is controlled by a social welfare maximizing fee, an observable queue will have higher welfare than the corresponding unobservable queue. In the observable case, a customer will enter only when it is socially desirable, whereas in the unobservable case only the joining probability is controlled and it is still possible for customers to enter when the queue is too long or to balk when it is too short. Hence, we get that $S^u_{soc} < S^o_{soc}$. Hassin proved the following results:

(1): If $R\mu \leq 2C$ then $S^u_m < S^o_m$ $\forall \lambda > 0$. This can be verified by comparing

$$S^u_m = \begin{cases} \left(\sqrt{R\mu} - \sqrt{C}\right)^2 & \text{when } \lambda \geq \mu - \sqrt{\frac{C\mu}{R}} \\ \lambda\left(R - \frac{C}{\mu-\lambda}\right) & \text{when } \lambda \leq \mu - \sqrt{\frac{C\mu}{R}} \end{cases} \tag{1.25}$$

.

with

$$S^o_m = \lambda R \frac{1-\rho^{n_m}}{1-\rho^{n_m+1}}\left(1 - \frac{n_m C}{\mu}\right).$$

Hence, the profit maximizer prefers to reveal the queue length to the customers if this can be done without cost.

(2): If $R\mu > 2C$ then a unique potential arrival λ^* exists so that

$$S_m^u > S_m^o \quad \text{for} \quad \lambda < \lambda*$$

, and

$$S_m^u \leq S_m^o \quad \text{for} \quad \lambda > \lambda* \,.$$

Thus, when $\lambda < \lambda*$ the profit maximizer prefers to conceal the queue length, whereas in the other case it prefers to disclose this information.

(3): The same properties apply to $S_{soc}^{o'}$ and S_{soc}^u as well for a different threshold λ_{**}. This can be verified by comparing

$$S_m^u = \begin{cases} \big(\sqrt{R\mu} - \sqrt{C}\big)^2 & \text{when } \lambda \geq \mu - \sqrt{\frac{C\mu}{R}} \\ \lambda\big(R - \frac{C}{\mu-\lambda}\big) & \text{when } \lambda \leq \mu - \sqrt{\frac{C\mu}{R}} \end{cases} \tag{1.26}$$

with

$$S_{soc}^{o'} = \lambda R\bigg(\frac{1-\rho^{n_m}}{1-\rho^{n_m+1}} - \frac{1}{V_e}\Big(\frac{1}{1-\rho} - \frac{\big(n_m+1\big)\rho^{n_m}}{1-\rho^{n_m+1}}\bigg).$$

Thus, when $\lambda < \lambda^{**}$ it is socially preferred that the profit maximizer does not reveal the queue length to the customers, whereas when $\lambda > \lambda_{**}$ an observable queue gives under profit maximization a higher value of social welfare than the observable queue.

(4): For arrival rates

$$\lambda_{**} < \lambda < \lambda_*,$$

the profit maximizer prefers to conceal the queue length. On the other hand, the social welfare would increase if the server could be induced to disclose this information.

(5): It is never socially worthwhile to induce the profit maximizer to conceal the queue length $\big($when he does not voluntarily do so$\big)$.

3.2 Models with delayed information structure

The effect of the information available to the customers on their strategic behavior and its economic considerations is a recurring theme in the thread of the queueing literature. This theory connects the observable and the unobservable cases.

In the models with delayed observation we suppose that there is a queueing system, where each customer makes his own decision without observing the system. These decisions may be 'joinor balk', 'stay or renege', 'to buy priority or not' etc. However, the customers later get informed about the number of customers in the queue. In the following chapter, we will present R. Hassin and Roet-Green's paper and Burnetas, Economou and Vasiliadis's paper discussing the influence of delayed information on the strategic customer's behavior.

3.2.1 The armchair decision: to depart towards the queue or not

Nowadays, technology offers more information to the public than ever. This affects the way service systems are modeled, since online information about congestion becomes more accessible to the customers, who can use it in their decisions. For example, hospitals in the USA publish their emergency rooms (ER) average waiting time on their websites (e.g. JFK medical center at jfkmc.com/our-services/er-wait-time.dot and Reston hospital center at www.restonhospital.com), International airports post online their security average waiting time information (e.g., Atlanta international airport in Georgia at www.atlanta-airport.com/ passenger /waittimes /default.aspx). In R.Hassin and Roet-Green's paper entitled 'The armchair decision: to depart towards the queue or not' is presented a model that allows the queue to evolve while the customer is on his way. Consider for instance a customer who wishes to go to a bank that offers the option to view its queue length online. The customer decides based on his inspection, whether to go to the bank or not, even though the queue length is likely to change during the customer's travel time. The expected changes are that the customers in the queue may be served or that others may arrive and join it. Yet, customers use the queue length information as an important variable that affects their decision on whether to depart or not. R.Hassin and Roet-Green refer to it as the armchair decision. Two main questions arise from this description. Firstly, how do customers take into consideration the changes in queue length during their travel time? Secondly, how will these changes affect customers' equilibrium strategy towards joining the queue? R.Hassin and Roet-Green presented in this paper various techniques in order to answer these questions. Specifically, they assumed that there is a system in which customers arrive according to a Poisson distribution with a parameter λ. The arriving customers can observe the current number of customers in the queue that they wish to join, but they do not know the number of customers that are on their way towards the queue. If a customer decides to depart and travel to the queue, the time he spends on his way is exponentially distributed with a parameter η. Moreover, they suppose that when a customer is in his way, he is not informed about any changes in the queue. The service is provided by a single server with exponentially service distribution with a param-

eter μ. A reward-cost structure in the system is also assumed. Specifically, each customer earns a reward R from his service completion and has a cost C_w per time unit he stays in the system. At the arrival moment, a customer observes the queue length and calculates his expected sojourn time depending on it. Consequently, according to the individual optimization in the observable case, each customer adopts a threshold strategy with Naor's threshold $n_e = \lfloor \frac{R\mu}{C_w} \rfloor$. Therefore, he will depart if his net benefit is not negative. The customers are assumed homogeneous. Each customer decides to depart towards the queue according to a strategy $\vec{P_D} = (P_D(1), P_D(2), \ldots, P_D(i), \ldots)$, where $P_D(i)$ is the probability to depart towards the queue if he observes i customers in it .
Furthermore, the assumption that

$$R > C_w(\frac{1}{\mu} + \frac{1}{\eta})$$

is being made, i.e. if an incoming customer believes that the system will be found empty, then his total reward R must be higher than his total cost $C_w(\frac{1}{\mu} + \frac{1}{\eta})$. Otherwise, the decision to balk will be a dominant strategy.

The queue state of the model is defined by the ordered pair $\big(i, j\big)$, where i is the number of customers in the queue and j is the number of customers that had already depart and are on their way to the system.

According to the model, the possible changes in the state space $\big(i, j\big)$ are the following:

$$\big(i, j\big) = \begin{cases} \big(i, j+1\big) & \text{if another customer departs, at rate } \lambda P_D\big(i\big) \\ \big(i-1, j\big) & \text{if a service completion occurs, at rate } \mu \\ \big(i+1, j-1\big) & \text{if a customer that was on his way joins the queue, at rate } j\eta \ . \end{cases}$$

However, in case that $i = n_e$ then the queue state shifts from state $\big(n_e, j\big)$ to state $\big(n_e, j-1\big)$ if a customer that was on his way arrives in the system. The following balance equations describe the evolution of the corresponding continuous-time Markov chain, of this model, for $i = 0, 1, \ldots, n_e$ and $j = 0, 1, 2, \ldots$:

$$\begin{aligned} \Big(\lambda p_D\big(i\big) + \mu\delta_i^o + \eta j\Big)\pi_{i,j} &= \lambda p_D\big(i\big)\delta_j^o\pi_{i,j-1} \\ &+ \mu\delta_i^{n_e}\pi_{i+1,j} + \eta\big(j+1\big)\Big(\delta_i^o\pi_{i-1,j+1} \\ &+ \big(1-\delta_i^{n_e}\big)\pi_{i,j+1}\Big), \end{aligned}$$

where:

$$\delta_i^o = \begin{cases} 1 & \text{when } \ i > 0 \\ 0 & \text{when } i = 0 \end{cases} \tag{1.27}$$

$$\delta_i^{n_e} = \begin{cases} 1 & \text{when } \ i < n_e \\ 0 & \text{when } i = n_e \end{cases} \tag{1.28}$$

Let $E(i, j)$ be the expected sojourn time for a tagged customer that departs if he could observe that the state is (i, j). Then

$$\begin{aligned} E\big(i, j\big) &= \frac{1}{\lambda P_D(i) + \mu\delta_i^o + \big(j+1\big)\eta} + \frac{\eta\delta_i^{n_e}}{\lambda P_D\big(i\big) + \mu\delta_i^o + \big(j+1\big)\eta} \cdot \frac{i+1}{\mu} \\ &+ \frac{j\eta}{\lambda P_D(i) + \mu\delta_i^o + \big(j+1\big)\eta} \cdot \big[E\big(i+1, j-1\big)\delta_i^{n_e} + E(i, j-1) \cdot (1 - \delta_i^{n_e}\big] \\ &+ \frac{\lambda P_D(i)}{\lambda P_D(i) + \mu\delta_i^o + (j+1)\eta} E\big(i, j+1\big) \\ &+ \frac{\mu\delta_i^o}{\lambda P_D(i) + \mu\delta_i^o + (j+1)\eta} \cdot E\big(i-1, j\big) \end{aligned}$$

The first term in the right part of the equation refers to the expected time until one of the following events happens first: a customer decides to depart towards the queue, a customer that was on his way arrived and a service completion occurs. The second term is the remaining expected sojourn time in the system for the tagged customer in case he is going to be the next to arrive and join the queue. The third term is the expected sojourn time for the tagged customer if one of the other j customers that are on their way arrives and joins the queue, or arrives and finds that the queue reached the threshold and therefore leaves the system. The fourth term is the expected sojourn time of the tagged customer if another customer departs towards the queue and the last term is the expected sojourn time if a service completion occurs.
The expected net benefit of a customer who observes i customers in the queue and decides to depart towards the queue is,

$$\begin{aligned} U\big(i\big) &= R - C_w E\big(i\big) = Rp - C_w \sum_{j=0}^{\infty} E\big(i, j\big) Pr\big(j|i\big) \\ &= R - C_w \sum_{j=0}^{\infty} E\big(i, j\big) \frac{\pi_{i,j}}{\pi_i}, \\ \pi_i &= \sum_{j=0}^{\infty} \pi_{ij} and pis the probability the i - tagged customer to receive his reward R. \end{aligned}$$

Thus, according to individual optimization, a customer will depart if $U\big(i\big) > 0$ thus, $P_D(i) = 1$ is best response, and balk if $U\big(i\big) < 0$ thus $P_D(i) = 0$ is best response.

In case that $U(i) = 0$, then the customer is indifferent between balking and departing towards the queue, so any $P_D(i) \in [0, 1]$ is best response. The system is characterized by three normalized parameters, namely $w = \frac{\eta}{\mu}$, which is the traveling parameter, $\rho = \frac{\lambda}{\mu}$, which is the congestion parameter, and $v = \frac{R\mu}{C_w}$, which is the normalized reward parameter. R.Hassin and Roet-Green have proved that for each set of normalized parameters ρ, w, v there is a symmetric Nash equilibrium strategy of this model. Specifically, they have proved the following result:

Theorem (3.1). *For each set of normalized parameters ρ, w, v there exists a symmetric Nash equilibrium strategy of this model.*

Proof. : A strategy in this game consists of a vector

$$P_D(x) = [P_D(0), P_D(1), \ldots, P_D(n_e)],$$

where $P_D(i)$ is the probability to depart towards the queue after observing i customers in it. If X is the space of mixed strategy vectors then, X is the n_e-dimensional cube, $X = [0, 1]^{n_e}$. Let $F : X \to X$ be the function that generates best response strategies:

$$F(x) = \{y \in X : y = P_D(x)\}, \quad \text{where} P_D(x) \in (0, 1).$$

F is a convex set of for each x. Indeed,
if $y_1, y_2 \in F(x)$, then $\forall w \in (0, 1)$, $y_3 = wy_1 + (1 - w)y_2 \in F(x)$.
In case that $y_1 = y_2$ then $y_3 = y_2 \in F(x)$. In case that $y_1 = y_2$ then we get that $y_3 = y_2 \in F(x)$. In case that $y_1 \neq y_2$ then at least one i exists so that $y_1(i) \neq y_2(i)$. Then, for every component i for which $y_1(i) \neq y_2(i)$ the customer is indifferent between departing and balking so y_3 is best response, therefore $y_3 \in F(x)$.Thus, F(x) is convex. Moreover, F is continuous as the composition of the steady-state probabilities and as the function which assigns the best response. Therefore, the graph of F is a closed set. According to Kakutani's fixed point theorem, F has a fixed point denoted by P_D^e. This strategy is best response of a player when all the other players use P_D^e strategy, which defines a symmetric Nash equilibrium strategy. □

R.Hassin and Roet -Green use an algorithm in order to compute the equilibrium strategy. Initially, they supposed that $j \leq N$ $(N >> n_e)$ and given the

expect sojourn time and the linear balance equations the algorithm is:
1. Choose an arbitrary strategy vector $[P_D]$, and define a tolerance parameter ϵ.

2.Compute the steady-state probabilities matrix $\left(\pi_{i,j}\right)$

3.Compute the expected sojourn time matrix $E\big(i,j\big)$ using the $\big(\pi_{i,j}\big)$ matrix.

4. Compute the utility vector U by using $\big(\Pi_{i,j}\big)$ and $E\big(i,j\big)$. Then for each $0 < i < n_e$:

a) If $U\big(i\big) > \epsilon$, set $P_D^*\big(i\big) = 1$

b) If $U\big(i\big) < -\epsilon$, set $P_D^*\big(i\big) = 0$

c) If $|U\big(i\big)| < \epsilon$, set $P_D^*\big(i\big) = P_D\big(i\big)$

5. While

$$\sum_{i=0}^{n_e} |P_D^*\big(i\big) - P_D\big(i\big)| \geq \epsilon,$$

define the new strategy$[P_D^{new}]$ as a convex combination of the old strategy $[P_D]$ and its best response $[P_D^*]$, using a random number $\gamma \in \big(0,1\big)$ as a weight. Continue from step 2.

6. If $\sum_{i=0}^{n_e} |P_D^* - P_D| < \epsilon$, then declare the equilibrium strategy as $P_D^e = P_D^*$.

If $\epsilon = 0,0005$ then the numerical results show that the algorithm always converges to an equilibrium strategy.
In the final chapter of their paper, Hassin and Roet-Green present some numerical results. According to the numerical study, the equilibrium seems to be sensitive to the initial strategy vector i.e. depending on the initial vector the type of the equilibrium differs. This results in multiple equilibrium strategies for many sets of normalized parameters, for example the equilibrium strategy may be a pure threshold strategy or a mixed threshold strategy i.e. a customer to depart if the number of customers is below the threshold, to balk if the number of customers in the queue is higher than the threshold and to waver between departing and balking when the queue length is equal to the threshold. Moreover, the equilibrium strategy may be a double threshold strategy, i.e. to depart when the queue length is below the first threshold, to balk if the queue length is between the first and the second threshold and to depart when the queue length is over the second threshold. Furthermore, according to numerical results the equilibrium strategy may not be a threshold strategy, i.e. the customers depart when the

queue length is below a specific number and waver between departing and balking when is over or equal to this specific number. However, under a specific set of parameters ρ, w and v the type of the equilibrium strategy is unique (equilibrium mixed threshold, double threshold, etc.) for any initial strategy $\vec{P_D}$. What is more, if $E_u = \sum_{i=0}^{n_e} \pi_i \max\left(U(i), 0\right)$ is the expected utility of the customers and P_{n_e} is the probability to reach its maximum length, then according to the numerical results for each set of parameters ρ, w, and v, the objectives E_u and Π_{ne} are received within a small range for all equilibrium strategies.

3.2.2 Strategic M/M/1 queueing model with synchronized delayed observations

Burnetas, Economou and Vasiliadis in their paper 'Strategic customer behavior in a queueing system considered that the customers arrive according to a Poisson process at rate λ at a service facility with infinite waiting space. There is a single server and the service time are exponentially distributed at rate μ. The queueing discipline is FCFS. Upon arrival, a customer decides whether to join or balk without observing the queue length. However,the administrator of the system announces to **al**l customers their positions in the system, at the points of a Poisson process at rate θ. Each customer, at his epochs $\big($arrival instant, announcement instant$\big)$ is supposed to know the operating parameters of the system, i.e. λ, μ, and θ. In addition, every customer anticipates a reward R from his service completion, whereas he accumulates waiting cost C per unit time for which he stays in the system.

The arrival, service and announcement processes are said to be independent. Moreover, retrials of balking customers are not permitted. When a customer arrives, the system is unobservable to him but at the epoch of the first announcement following his entrance the system becomes observable to him. Because of the exponential assumptions, the customers may renege only at announcements instants. Additionally, because of the FCFS queue discipline and the full observation structure a customer has a possible incentive to renege only at the first announcement. The strategic behavior of a customer regarding joinning/balking is specified by a joining probability q_*. A customer stays after the first announcement, if his position n at the system is such that his net benefit $(R - C\frac{n}{\mu})$ is positive. Thus, each customer has a reneging threshold n_*, which determines if a customer will remain in the system or not, at the time of the first announcement after their arrival. Thus, the strategy of an incoming customer is $\big(n_*, q\big)$. Under such a strategy the number of customers in the system is represented by a continuous time Markov chain $N(t)$ with transition rates

$$q_{i,j} = \begin{cases} \lambda q_* & \text{when } i \geq 0, j = i+1, \\ \mu & \text{when } i \geq 1, i \neq n_* + 1, j = i-1, \\ \theta & \text{when } i \geq n_* + 2, j = n_*, \\ \mu + \theta & \text{when } i = n_* + 1, j = n_*, \\ 0 & \text{otherwise} \end{cases} \tag{1.29}$$

The stationary distribution of $\{N(t)\}$ and its mean are given in the following proposition.

Proposition 3.1 The stationary distribution of the number of customers in the system $N(t)$ when the customers follow an (n_*, q_*) strategy, with $n_*, q_* > 0$ and $\lambda q_* \neq \mu$ is given by

$$\pi(n_*, q) = \begin{cases} B_* \rho_{*1}^{n} & \text{when } 0 \leq n \leq n_* - 1 \\ \\ B_* \rho_{*1}^{n_*} \rho_{*2}^{n-n_*} & \text{when } n \geq n_* \end{cases} \tag{1.30}$$

where,

$$\rho_{*1} = \frac{\lambda q_*}{\mu}, \rho_{*2} = \frac{\lambda q_* + \mu + \theta - \sqrt{(\lambda q_* + \mu + \theta)^2 - 4\lambda q_* \mu}}{2\mu}$$

and,

$$B_* = \frac{(1-\rho_{*1})(1-\rho_{*2})}{1-\rho_{*2}-\rho_{*1}^{n_*+1}+\rho_{*1}^{n_*}\rho_{*2}}.$$

The corresponding mean stationary number of customers in the system is

$$E_{(n_*,q_*)}(N) = \frac{(1-\rho_{*2})[(n_*-1)\rho_{*1}^{n_*+1} - n_*\rho_{*1}^{n_*}}{+\rho_{*1}}(1-\rho_{*1})[1-\rho_{*2}-\rho_{*1}^{n_*+1}+\rho_{*1}^{n_*}\rho_{*2}] + \frac{(1-\rho_*)[n_*\rho_{*1} - (n_*-1)\rho_{*1}^{n_*}\rho_{*2}]}{(1-\rho_{*2})[1-\rho_{*2}-\rho_{*1}^{n_*+1}+\rho_{*1}^{n_*}\rho_{*2}]}$$

In the next proposition the expected net benefit of a customer that arrives when there are n customers in the system (excluding himself) and decides to join is represented.

Proposition 3.2 Consider the $M/M/1$ queue with delayed observations, where the customers follow an (n_*, q_*) strategy. Then, the conditional expected net benefit of a customer that arrives when there are n in the system and decides to join and uses the same reneging threshold n_* as the other customers is given by

$$U(n|n_*) = \begin{cases} R - C\frac{n+1}{\mu} & \text{when } 0 \leq n \leq n_* - 1 \\ (R - \frac{Cn_*}{\mu} + \frac{C}{\theta})(\frac{\mu}{\mu+\theta})^{n-n_*+1} - \frac{C}{\theta} & \text{when } n \geq n_* \end{cases} \quad (1.31)$$

Moreover, the expected net benefit of a customer that decides to join is,

$$\begin{aligned} U(n_*, q_*) &= B_*\Big(R - \frac{C}{\mu}\Big)\frac{1-\rho_{*1}^{n_*}}{1-\rho_{*1}} - B_*\frac{C}{\mu}\frac{(n_*-1)\rho_{*1}^{n_*+1} - n_*\rho_{*1}^{n_*} + \rho_{*1}}{(1-\rho_{*1})^2} \\ &+ B_*(R - \frac{Cn_*}{\mu} + \frac{C}{\theta})\frac{\mu\rho_{*1}^{n_*}}{\mu+\theta-\mu\rho_{*2}} - B_*\frac{C}{\theta}\frac{\rho_{*1}^{n_*}}{1-\rho_{*2}} \end{aligned}$$

Proof Consider a tagged customer that arrives and decides to join, when the system has n other customers.

Case I: $n \geq n_* - 1$**.** In this case the customer will receive his reward R, since he has no incentive to renege. Additionally, his mean waiting time in the system provided that there are $n-1$ customers in the system is $\frac{n+1}{\mu}$. Thus, we get the first branch of the equation (1.32).

Case II: $n > n_*$**.** In the second case the net benefit of the tagged customer U_n has the representation

$$U_n = (R - C(Y_n + Z))1_{\{X \geq Y_n\}} - CX1_{\{X<Y_n\}},$$

where$X$, Y_n and Z are independent random variables. X is an $Exp(\theta)$ distributed random variable, which represents the time till the first announcement after the arrival of the tagged customer, Y_n is an $Erlang(n - n_* + 1, \mu)$ random variable and represents the time till the tagged customer has no incentive to renege and Z is an $Erlang(n_*, \mu)$ random variable, which represents the time after Y_n till the departure of the tagged customer (if he does not renege). In case that the first announcement after the arrival of the tagged customer happens before the time till the tagged customer has no incentive to renege then his sojourn time is X. Otherwise, the customer will stay in the system for $Y_n + Z$ time units and will receive the reward for service. Therefore, taking expected values $U(n|n_*) = (R - CE(Z))P[X \geq Y_n] - CE[\min(X, Y_n]$. By computing $P[X \geq Y]$ and $E[\min(X, Y_n)]$ we get the computation of the expected net benefit. Moreover, by evaluating the geometric sums we can determine the expected net benefit by the formula $U(n_*, q_*) = \sum_{n=0}^{\infty} \pi_n(n_*, q_*)U(n|n_*)$.

Burnetas, Economou and Vasiliadis reported several monotonicity results that concern the process $N(t)$ that records the numbers of customers in the system, the throughput of the system, the conditional expected net benefit $U(n|n_*)$ and the expected net benefit $U(n_*, q_*)$. Specifically, the following monotonicity results are valid:

- By using sample path arguments they proved that $N(t)$ is:

  - Stochastically increasing in the joining rate λq_* when the other parameters are kept fixed,

  - Stochastically decreasing in the service rate μ when the other parameters are kept fixed,

  - Stochastically decreasing in the announcement rate θ when the other parameters are kept fixed,

  - Stochastically increasing in the reneging threshold n_* when the other parameters are kept fixed.

- The throughput of the system is a decreasing function of θ, for any fixed strategy (n_*, q_*) of the customers and the other parameters $\big(\lambda, \mu, R, \text{and } C\big)$ fixed.

- The conditional expected net benefit $U(n|n_*)$ is a strictly decreasing function of n, for any fixed strategy
$n_* \leq \frac{R\mu}{C} + \frac{\mu}{\theta}$,

- When the customers follow an (n_*, q_*) strategy then the expected net benefit $U(n_*, q_*)$ is a strictly decreasing function of q_* for any fixed $n_* \leq \frac{R\mu}{C} + \frac{\mu}{\theta}$

By using the monotonicity and continuity of $U(n_*, q_*)$, in q_*, the fact that the best strategy of a tagged customer is to stay if his position n at the first announcement is such that $n \leq n_e$, with $n_e = \lfloor \frac{\mu R}{C} \rfloor$, and that $n_e \leq \frac{R\mu}{C} + \frac{\mu}{\theta}$, Burnetas, Economou and Vasiliadis proved that in the $M/M/1$ with delayed observations an equilibrium strategy always exists and it is unique. Specifically, we have the following theorem.

Theorem (3.2)**.** *In the $M/M/1$ queue with synchronized delayed observations an equilibrium strategy always exists and is unique. If n_e is Naor's threshold then:*

Case I: *$U(n_e, 0) \leq 0$. Then, $(n_e, 0)$ is the unique equilibrium strategy.*

Case II: *$U(n_e, 1) < 0 < U(n_e, 0)$. Then, the equation $U(n_e, q_*) = 0$ has exactly one root with respect to q_* in $(0, 1)$ and (n_e, q_e) is the unique equilibrium strategy.*

Case III: *$U(n_e, 1) \geq 0$. Then, $(n_e, 1)$ is the unique equilibrium strategy.*

In this paper the effects of the announcement rate θ on the equilibrium join probability q_e, the equilibrium throughput and the equilibrium social welfare, when all other λ, μ, R and C are kept fixed are also analyzed.

By using that the expected net benefit $U(n_e, q_*)$ is a strictly increasing function of θ, for fixed other parameters λ, μ, R, C, $q_* \neq 0$ and $n_e = \lfloor \frac{\mu R}{C} \rfloor$, it is proved that the equilibrium join probability is an increasing function of θ. More concretely, they proved the following theorem:

Theorem (3.3)**.** *In the $M/M/1$ queue with delayed observations, the unique equilibrium join probability $q_e(\theta)$ is an increasing function of the announcement rate θ, when all other parameters of the model λ, μ, R and C are kept fixed. In particular, when $\lambda < \mu$, we have the following cases as R varies in $(0, \infty)$.*

***Case I**: $R \leq \frac{C}{\mu}$. Then, $q_e(\theta) = 0$, for $\theta \in (0, \infty)$.*

***Case II**: $\frac{C}{\mu} < R < \frac{C}{\mu - \lambda}$. Then,*

$$\lim_{\theta \to \ 0} q_e(\theta) = \frac{1}{\lambda}(\mu - \frac{C}{R} \in (0, 1)$$

and $q_e(\theta)$ is increasing for $\theta \in (0, \infty)$. Moreover, there exists a $\theta_0 > 0$ such that $q_e(\theta) < 1$ for $\theta \in (0, \theta_0)$, while $q_e(\theta) = 1$ for $\theta \in [\theta_0, \infty)$.

***Case III**: $R \geq \frac{C}{\mu - \lambda}$. Then, $q_e(\theta) = 1$, for $\theta \in (0, \infty)$.*

When $\lambda \geq \mu$, Case I is still valid as above, but Case II corresponds to $R > \frac{C}{\mu}$, while Case III does not exist.

The previous Theorem implies the following proposition for the equilibrium throughput:

Proposition 3.3 The equilibrium throughput, π_0^e is a constant zero function of θ when $R \leq \frac{C}{\mu}$, while it is a decreasing function of θ in case that $R \geq \frac{C}{\mu-\lambda}$.

Proof For $R \leq \frac{C}{\mu}$, according to the previous theorem we have that $q_e(\theta) = 0$, thus, $\pi_0 = 0 \ \forall\theta \in (0, \infty)$. Moreover, for $R \geq \frac{C}{\mu-\lambda}$ we get that $q_e(\theta) = 1$ $\frac{\partial \pi_0^e}{\partial \theta} = \frac{\partial \pi_0}{\partial \theta} + \frac{\partial \pi_0}{\partial q_e}\frac{\partial q_e}{\partial \theta} = \frac{\partial \pi_0}{\partial \theta} < 0$, because we have mentioned that the throughput of the system is a decreasing function of θ. Thus, the equilibrium throughput is a decreasing function of θ as well.

Remark The aforementioned monotonicity of the equilibrium throughput is still valid when $\frac{C}{\mu} < R < \frac{C}{\mu-\lambda}$ when $\theta \geq \theta_0$, i.e, the throughout is decreasing in θ. There is also extensive numerical evidence that for $\frac{C}{\mu} < R < \frac{C}{\mu-\lambda}$ the equilibrium throughput is a unimodal function of θ and the mode occurs in θ_0.

Let, $S_e = \lambda q_e U(n_e, q_e)$ be the equilibrium social welfare. We consider S_e as a function of the announcement rate θ. We have showed that the $U(n_e, q_e) = 0$ $\forall q_e \in (0, 1)$ thus, $S_e = 0 \ \forall q_e \in [0, 1)$. On the other hand, the equilibrium join probability $q_e(\theta)$ is an increasing function of θ when the other parameters are kept fixed and $U(n_e, 1)$ is an increasing function of θ. Therefore, when $q_e(\theta)$ reaches 1 the social welfare function becomes a strictly increasing function of θ. Therefore, we conclude that S_e is a non-decreasing function of θ, which is 0 for $\theta < \theta_0$.

In the case of social optimization the corresponding function is too complicated to be maximized so that the optimal join probability and the optimal threshold cannot be determined in closed-form expressions. Of course, an incoming customer brings a burden to the other customers by increasing the overall delay, so $n_{soc} \leq n_e$. However, the order of q_{soc} and q_e is less clear. Specifically, an increase on the join probability has two opposite effects on the social welfare: On the one hand, it may increase the expected reward from service completions, but on the other hand incurs greater waiting costs. In case that the reneging threshold is low and the announcement rate is high then the expected waiting cost can be kept low, even if the administrator uses a higher join probability. Thus, $q_{soc} > q_e$ may be valid in contrast to the results in the unobservable model. Moreover, numerical results also verify that the social administrator may sometimes prefer to accept more customers than those who would be wiling to enter following an individual equilibrium strategy. This shows that the system with the delayed announcements gives extra flexibility regarding the maximization of the social welfare. Numerical results also show that the system throughput is always lower under the socially optimal strategy than under the equilibrium strategy. In the paper it is also analyzed whether a socially optimal strategy can be induced by imposing appropriate fees. They refer to this property as coordination. Specifically, by imposing an entrance fee f_e that is paid upon joining in the system, the administrator of the system has the possibility to influence the customers. Then,the equilibrium reneging threshold is still n_e and the expected net benefit of a customer who decides to join when the others join with probability q is

$$U_e\big(q, f_e\big) = U\big(n_e, q\big) - f_e.$$

On the other hand, if the administrator of the system sets a service fee f_s that is paid only if a customer gets served, then the equilibrium threshold becomes

$$n_e\big(f_s\big) = \left\lfloor \frac{\mu\big(R - f_s\big)}{C} \right\rfloor$$

and the the expected net benefit of a customer is given by

$$U_s\big(q, , f_s\big) = U\big(n_e\big(f_s\big), q, R - f_s\big).$$

It also possible that the administrator imposes both entrance fees and service fees. Thereafter, $U_{es}(q; f_e, f_s) = U(n_e(f_s), q, R - f_s) - f_e$. Furthermore, in case that coordination is not possible, it would be important to find the discrepancy between the socially optimal strategy and the best achievable strategy by imposing either entrance fee or service fee or both of them. The most significant robust insights obtained by numerical results are the following:

- The use of the entrance fee only rarely leads to the coordination of the system and the achievable fraction of the optimal social benefit varies.

- Imposing only the service fee in most cases is sufficient for the coordination or the achievement of a high fraction of the optimal social benefit.

4 An M/M/1 model with independent delayed observations

Model description
In this chapter we would like to study the influence of the information available to the customers, on their strategic behavior. Specifically, we will present a queueing model which has the same assumptions as the

$$(M/M/1 \text{queueing model with delayed observations})$$

which is presented in Burnetas, Economou and Vasiliadis's paper, but in this case the difference is that the administrator later announces to each customer, independently, his current position instead of announcing to all the customers simultaneously their current positions.
Consider for example a bank that offers the option to view its queue length online. A customer who applies in order to get a service in the bank after a specific time period receives an e-mail about the number of his application. As a result, he knows the number of customers that are in front of him in the queue by obser ving the queue length online. Our goal is to determine in what way this delayed information can influence a customer's behavior. We are attempting to answer this question in the following project.
More concretely, we assume an $M/M/1$ queueing system in which customers arrive according to a Poisson process at a rate λ and the service times are exponentially distributed at rate μ. Moreover, we assume that there is an administrative manager that announces to each customer independently an exponentially distributed time with parameter θ after his arrival. Each customer earns a known reward R from his service completion, whereas he has a waiting cost at a known rate C as long as he stays in the system. After the announcement of his position a customer reevaluates his expected welfare of staying in the system and he will renege if the expected benefit is negative. The arrival, service and announcement processes are assumed to be independent and we assume that retrials of balking customers are not permitted. We are interested in the study of the strategic behavior of the customers, regarding their joining/balking and staying/reneging dilemmas, in this system.

We assume that the customers are homogeneous. Therefore the service completion reward and the cost per time unit staying in the system are the same for all customers. Therefore, we suppose that all the players follow the same strategy. According to this assumption we are interested in finding only symmetric best responses.
We consider that each customer's strategy is a vector $\big(q,\, n_*\big)$ that belongs in the product space $[0,1] \times \mathbb{N}$. The first component of the vector, (q), represents the join probability of a customer under the assumption that he is not able to observe the queueing length and the second component of the vector, (n_*), represents the reneging threshold at the epoch the system becomes observable to a customer. Below, we will study the behavior of the customers in the system, when they all use the strategy $\big(q,\, n_*\big)$.

Let $N(t)$ be the stochastic process that represents the number of customers in the system at time t. Then the transitions of this Markov chain are.

$$q_{ij} = \begin{cases} \lambda q_* & \text{when } i \geq 0, j = i+1 \\ \mu & \text{ when } 1 \leq i \leq n_* \ , \ j = i-1 \\ \mu + \big(i - n_*\big)\theta & \text{when } i \geq n_* + 1 \ , \ j = i-1 \\ 0 & \text{otherwise} \end{cases} \tag{1.32}$$

The transition rate diagram of $N(t)$ is given in figure 1.

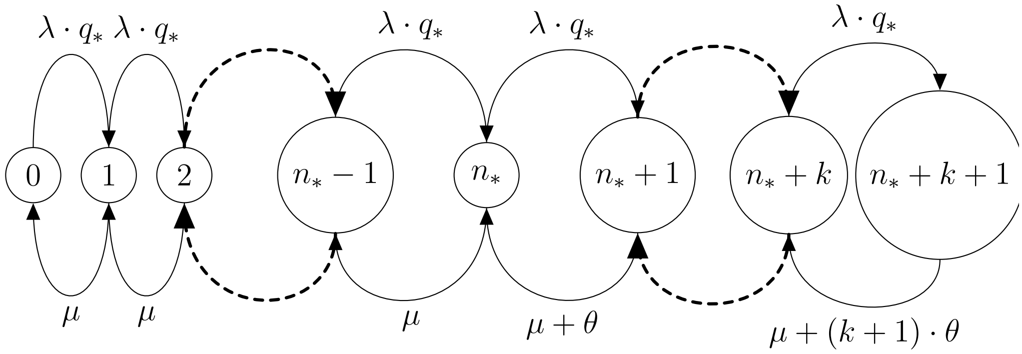


Figure1: Transition rate diagram of $N(t)$

Proposition 4.1: The stationary distribution of the number of customers in the queueing system under the strategy $\left(n_*, q\right)$ and under the utilization restriction $\left(\lambda q_*\right) < \mu$ is given by the following formula,

$$\pi\left(n_*, q\right) = \begin{cases} k_1^i \pi_0 & \text{when } i \leq n_* \\ & \\ \frac{k_3^{i-n_*}\Gamma\left(k_2\right)}{\Gamma\left(k_2+i-n_*\right)} k_1^{n_*} \pi_0 & \text{when } i \geq n_* + 1 \end{cases} \tag{1.33}$$

where

$$k_1 = \frac{\lambda q_*}{\mu}, \quad k_2 = \frac{\mu}{\theta}, k_3 = \frac{\lambda q_*}{\theta},$$

$$\pi_0 = \frac{1}{A},$$

$$A = \frac{1-k_1^{n_*+1}}{1-k_1} + k_1^{n_*}\Gamma\big(k_2\big)\Big(E_{1,k_2}\big(k_3\big) - \frac{1}{\Gamma\big(k_2\big)}\Big)$$

and

$$E_{1,k_2}\big(k_3\big) = \sum_{n=0}^{\infty} \frac{k_3^n}{\Gamma\big(n+k_2\big)}$$

the Mittag-Leffler function.

Proof. : As the stochastic model is a birth-and-death process we conclude that the process is reversible. Therefore if S is the state space, we get that for every $i, j \in S$,

$$\pi_i q_{ij} = \pi_j q_{ji}.$$

Specifically, if $i = 1, 2, \ldots, n_* - 1$ we get that,

$$\begin{aligned} \lambda q_* \pi_i &= \mu \pi_{i+1} \iff \\ \pi_i &= \Big(\frac{\lambda q_*}{\mu}\Big)^i \pi_0 \quad \forall i \leq n_*. \end{aligned} \tag{1.34}$$

If $i \geq n_*$ then we get the following equation

$$\begin{aligned} \pi_i \lambda q_* &= \Big(\mu + \big(i+1-n_*\big)\theta\Big)\pi_{i+1} \iff \\ \pi_{i+1} &= \frac{\lambda q_*}{\Big(\mu + \big(i+1-n_*\big)\theta\Big)}\pi_i. \end{aligned} \tag{1.35}$$

Equations $\big(1.35\big)$ and $\big(1.36\big)$ yield

$$\begin{aligned} \pi_{n_*+1} &= \frac{\lambda q_*}{\mu+\theta}\pi_{n_*} \iff \\ \pi_{n_*+1} &= \frac{\lambda q_*}{\mu+\theta}\big(\frac{\lambda q_*}{\mu}\big)^{n_*}\pi_0. \end{aligned} \tag{1.36}$$

Equations $\big(1.36\big)$ and $\big(1.37\big)$ yield

$$\pi_i = \frac{\Big(\lambda q_*\Big)^{i-n_*}}{\prod_{j=1}^{i-n_*}\big(\mu + j\theta\big)}\Big(\frac{\lambda q_*}{\mu}\Big)^{n_*}\pi_0 \quad \forall i \geq n_*.$$

By using the normalizing equation we get that

$$\sum_{i=0}^{\infty}\pi_i = 1 \iff$$

$$\sum_{i=0}^{n_*}\Big(\frac{\lambda q_*}{\mu}\Big)^i\pi_0 + \sum_{i=n_*+1}^{\infty}\frac{\left(\lambda q_*\right)^{i-n_*}}{\prod_{j=1}^{i-n_*}\left(\mu+j\theta\right)}\Big(\frac{\lambda q_*}{\mu}\Big)^{n_*}\pi_0 = 1. \tag{1.37}$$

We will try to give an analytical formula for the sum

$$\sum_{i=n_*+1}^{\infty}\frac{\left(\lambda q_*\right)^{i-n_*}}{\prod_{j=1}^{i-n_*}\left(\mu+j\theta\right)}.$$

If we denote $k=i-n_*$, then

$$\sum_{k=1}^{\infty}\frac{\left(\lambda q_*\right)^k}{\left(\mu+\theta\right)\left(\mu+2\theta\right)\cdots\left(\mu+k\theta\right)} = \sum_{k=1}^{\infty}\frac{\left(\lambda q_*\right)^k}{\theta^k\left(\frac{\mu}{\theta}+1\right)\left(\frac{\mu}{\theta}+2\right)\cdots\left(\frac{\mu}{\theta}+k\right)}. \tag{1.38}$$

If we use the equation

$$\frac{\Gamma\left(a\right)}{\Gamma\left(a-1\right)} = a-1$$

and the Pochhammer symbol

$$\frac{\Gamma\left(\frac{\mu}{\theta}+n\right)}{\Gamma\left(\frac{\mu}{\theta}\right)} = \Big(\frac{\mu}{\theta}\Big)_k,$$

according to $\left(1.39\right)$ we get that

$$\sum_{k=1}^{\infty}\Big(\frac{\lambda q_*}{\theta}\Big)^k\frac{1}{\left(\frac{\mu}{\theta}\right)_k} = \sum_{k=1}^{\infty}\frac{\Gamma\left(\frac{\mu}{\theta}\right)}{\Gamma\left(\frac{\mu}{\theta}+k\right)}\Big(\frac{\lambda q_*}{\theta}\Big)^k = \Gamma\Big(\frac{\mu}{\theta}\Big)\left(E_{1,\frac{\mu}{\theta}}\Big(\frac{\lambda q_*}{\theta}\Big) - \frac{1}{\Gamma\left(\frac{\mu}{\theta}\right)}\right). \tag{1.39}$$

Thus, we proved that $\sum_{i=n_*+1}^{\infty} \frac{\left(\lambda q_*\right)^{i-n_*}}{\prod_{j=1}^{i-n_*}\left(\mu+j\theta\right)} = \Gamma\left(\frac{\mu}{\theta}\right)\left(E_{1,\frac{\mu}{\theta}}\left(\frac{\lambda q_*}{\theta}\right) - \frac{1}{\Gamma\left(\frac{\mu}{\theta}\right)}\right)$. If we denote

$$k_1 = \frac{\lambda q_*}{\mu}, \quad k_2 = \frac{\mu}{\theta}, \quad k_3 = \frac{\lambda q_*}{\theta}$$

and if we use the equation

$$\sum_{i=0}^{n_*} k_1^i = \frac{1-k_1^{n_*+1}}{1-k_1}$$

then equation $\left(1.38\right)$ yields

$$\pi_0\Big(\frac{1-k_1^{n_*+1}}{1-k_1} + k_1^{n_*}\Gamma\left(k_2\right)\left(E_{1,k_2}\left(k_3\right) - \frac{1}{\Gamma\left(k_2\right)}\right)\Big) = 1.$$

Thus,

$$\pi_0 = \frac{1}{\left(\frac{1-k_1^{n_*+1}}{1-k_1} + k_1^{n_*}\Gamma\left(k_2\right)\left(E_{1,k_2}\left(k_3\right) - \frac{1}{\Gamma\left(k_2\right)}\right)\right)}.$$

□

Proposition 4.2: The mean stationary number of customers in the system $E_{\left(n_*,q\right)}\left(N\right)$ under a specific strategy (n_*, q) is given by the formula

$$
\begin{aligned}
E_{\left(n_*,q\right)}\left(N\right) &= \pi_0\Bigg[\frac{-\left(n_*+1\right)k_1^{n_*+1}\left(1-k_1\right)+\left(1-k_1\right)^{n_*+1}k_1}{\left(1-k_1\right)^2} \\
&+ k_1^{n_*}\Gamma\left(k_2\right)\left(n_*\left(E_{1,k_2}\left(k_3\right)\right)-\frac{1}{\Gamma\left(k_2\right)}\right) \\
&+ \Psi_{2,2}\left[\begin{array}{c}\left(1,1\right)\left(1,1\right) \\ \left(1,0\right),\left(1,k_2\right)\end{array};k_3\right]-\frac{1}{\Gamma\left(k_2\right)}\Bigg]
\end{aligned}
$$

$$
\text{where}\ \ k_1 = \frac{\lambda q_*}{\mu}, k_2 = \frac{\mu}{\theta},\quad k_3 = \frac{\lambda q_*}{\theta}\ \ \text{and}
$$

$$
\Psi_{2,2}\left[\begin{array}{c}\left(1,1\right)\left(1,1\right) \\ \left(1,0\right),\left(1,k_2\right)\end{array};k_3\right] = \sum_{k=0}^{\infty}\frac{\Gamma\left(k+1\right)\Gamma\left(k+1\right)}{\Gamma\left(k\right)\Gamma\left(k+2\right)}\frac{k_3^k}{k!}\ \ \text{the Fox-Wright function.}
$$

Proof.

$$E_{\left(n_*,q\right)}\left(N\right) = \sum_{i=0}^{\infty} i\pi_i = \sum_{i=0}^{n_*} ik_1^i\pi_0 + \sum_{i=n_*+1}^{\infty} i\frac{\left(k_3\right)^{i-n_*}\Gamma\left(k_2\right)}{\Gamma\left(k_2+i-n_*\right)}k_1^{n_*}\pi_0$$

$$= \pi_0\left(k_1\frac{\partial\left(\sum_{i=0}^{n*} k_1^i\right)}{\partial k_1} + k_1^{n_*}\Gamma\left(k_2\right)\sum_{k=1}^{\infty}\left(k+n_*\right)\frac{k_3^k}{\Gamma\left(k+k_2\right)}\right)$$

$$= \pi_0\left(k_1\left(\frac{-\left(n_*+1\right)k_1^{n_*}\left(1-k_1\right)+\left(1-k_1\right)^{n_*+1}}{\left(1-k_1\right)^2}\right)\right.$$

$$+ \left.k_1^{n_*}\Gamma\left(k_2\right)\left(n_*\sum_{k=1}^{\infty}\frac{\left(k_3\right)^k}{\Gamma\left(k+k_2\right)} + \sum_{k=1}^{\infty}\frac{\Gamma\left(k+1\right)\Gamma\left(k+1\right)}{\Gamma\left(k\right)\Gamma\left(k+k_2\right)}\frac{\left(k_3\right)^k}{k!}\right)\right)$$

$$= \pi_0\left[k_1\left(\frac{-\left(n_*+1\right)k_1^{n_*}\left(1-k_1\right)+\left(1-k_1\right)^{n_*+1}}{\left(1-k_1\right)^2}\right)\right.$$

$$+ \left.k_1^{n_*}\Gamma\left(k_2\right)\left(n_*\left(E_{1,k_2}\left(k_3\right)-\frac{1}{\Gamma\left(k_2\right)}\right)+\Psi_{2,2}\left[\begin{array}{c}\left(1,1\right)\left(1,1\right)\right)\\ \left(1,0\right),\left(1,k_2\right)\end{array};k_3\right]-\frac{1}{\Gamma\left(k_2\right)}\right)\right].$$

$\square$

Proposition 4.3: If all customers follow the (n_*, q) strategy, then the conditional expected net benefit for a tagged customer, who arrives when there are $n-1$ customers in the system, decides to join and uses the same reneging threshold n_* is given by the formula

$$U(n|n_*) = \begin{cases} R - C\frac{n}{\mu} & \text{when } 1 \le n \le n_* \\ R\frac{\mu}{\mu+(n-n_*)\theta} - C\frac{n}{\mu+(n-n_*)\theta} & \text{when } n \ge n_* + 1 \end{cases} \tag{1.40}$$

.

Proof. : We consider a tagged customer that arrives when there are $n-1$ customers and decides to join the system. We consider two cases according to whether $n \ge n_*$ or not

CASE I: If $1 \le n \le n_*$. Then the customer will receive the reward almost surely since he is not going to renege at the announcement of his position. Furthermore, when the position of a customer is n, then his sojourn time is an Erlang-distributed random variable with mean $\frac{n}{\mu}$. According to the Markovian property we conclude that the remaining time of the customer in service is $Exp(\mu)$. Therefore, the expected total waiting cost of the tagged customer is

$$C \cdot \frac{n}{\mu}$$

and the first branch of the equation follows.

CASE II: $n \ge n_* + 1$.
In order to find $U(n|n_*)$, we need to find the mean sojourn time of the tagged customer in the system. However, in the second case there is a positive probability the customer to renege without completing his service. Therefore, in order to find his conditional expected net benefit, we have also to determine the probability of getting his reward.
We note that the probability of the customer getting the reward is the same with the probability the customer to arrive at position n_*, before the announcement of his position. We use the first step analysis. We will find the probability of the position of the tagged customer becoming the $(n_* - 1)$-position before the announcement of his position. As the position of the customer is n, then there

are $n-n_*-1$ customers in front of him who are waiting in the queue, until the threshold n_*. Therefore, there are $n-n_*$ possible events that can change the position of the n-tagged customer from n to $n-1$. Equivalently, there are $n-n_*$ exponentially distributed random variables which represent the moments that each of these events happens. Specifically, we suppose that Y_1 stands for the time until a service completion which is $\text{Exp}\big(\mu\big)$ distributed and X_j represents the time until the j-th customer learns his position, or equivalently the announcement of his position occurs, $j \in \big(n_*+1, n_*+2, \cdots, n\big)$ which is $\text{Exp}\big(\theta\big)$ distributed. We mention that if

$$W_1, \cdots, W_n \backsim Exp\big(\lambda_n\big)$$

then

$$P\big(W_i = \min\big(W_1, \cdots, W_n\big)\big) = \frac{\lambda_i}{\sum_{i=1}^{n} \lambda_i}. \tag{1.41}$$

The equation (1.41) immediately yields that

$$P\Big(Y_1 = \min\big(Y_1, X_{n_*}, X_{n_*+1}, \cdots, X_n\big)\Big) = \frac{\mu}{\mu + \big(n-n_*\big)\theta}$$

,

$$P\Big(X_j = \min\big(X_{n_*}, X_{n_*+1}, \cdots, X_n\big)\Big) = \frac{\theta}{\mu + \big(n-n_*\big)\theta}.$$

In case that the minimum of the exponential times occurs in some X_n, then the tagged customer will learn his position, which is above his reneging threshold n_* and as a result he will abandon the system without receiving his reward, so the probability of getting his reward is 0. So we get that

$$P\big(\text{the tagged customer receive his reward}\big) =$$

$$\Big(\frac{\mu + \big(n-n_*-1\big)\theta}{\big(n-n_*\big)\theta\big)}\Big) P\Big(\text{the } \big(\text{n-1}\big)\text{-tagged customer receive his reward}\Big). \tag{1.42}$$

For $n = n_* + 1$ we get that

$$
\begin{aligned}
P\Big(\big(n_*+1\big)\text{-tagged customer receive his reward}\Big) &= \\
\frac{\mu}{\mu+\theta} P\Big(\text{the } n_* \text{ customer get his reward}\Big) &= \\
\frac{\mu}{\mu+\theta} \cdot 1 &= \frac{\mu}{\mu+\theta} \qquad (1.43)
\end{aligned}
$$

The probability that $\big(n_*+2\big)$-player gets his reward is:

$$
\begin{aligned}
P\Big(\big(n_*+2\big)\text{-player get the reward}\Big) &= \\
P\Big(\text{the } \big(n_*+2\big)\text{-player becomes the } \big(n_*+1\big)\text{-player}\Big) &\cdot \\
\cdot P\Big(\text{the } \big(n_*+1\big)\text{-player get his reward}\Big) &= \\
\frac{\mu+\theta}{\mu+2\theta}\frac{\mu}{\mu+\theta} &= \frac{\mu}{\mu+2\theta}.
\end{aligned}
$$

We suppose that the equation $\big(1.42\big)$ holds for

$$n = n_* + k$$

and we will show that is also true for

$$n = n_* + k + 1.$$

Indeed,

$$P\Big(\text{the } \big(n_* + k + 1\big)\text{-customer get the reward}\Big) \quad =$$

$$P\Big(\text{the } \big(n_* + k + 1\big)\text{-customer becomes the } \big(n_* + k\big)\text{-customer}\Big) \quad \cdot$$

$$\cdot P\Big(\text{the } \big(n_* + k\big)\text{-customer get the reward}\Big) \quad =$$

$$\frac{\mu + k\theta}{\mu + \big(k+1\big)\theta}.$$

Therefore by using induction on n we find the probability of a customer getting his reward. In order to prove that the mean sojourn time of the n-th customer in the system is

$$\frac{n}{\mu + \big(n - n_*\big)\theta}$$

we will use again the first step analysis and by induction in n we will get the result. The $n = n_* + 1$ -customer will stay in the system until the first event of either a service completion or his position announcement happens or equivalently $\frac{1}{\mu+\theta}$. Additionally, he will stay the mean sojourn time of the $n_* - th$customer if a service completion happens before his position announcement and in case that he learns his position before a service completion happens then he will leave the system immediately. We notice that the sojourn time of the n_* customer is a $\Gamma\big(n_*, \mu\big)$ distributed random variable, therefore, the mean sojourn time of the

n_*-customer in $\frac{n_*}{\mu}$. Thus, we have that

$$E\Big(\text{sojourn time of the } n_*\text{- customer}\Big) \quad =$$

$$\frac{1}{\mu+\theta} + \frac{\mu}{\mu+\theta}\cdot 0 + \frac{\mu}{\mu+\theta}E\Big(\text{sojourn time of } n_*\text{-customer}\Big) \quad =$$

$$\frac{1}{\mu+\theta} + \frac{\mu}{\mu+\theta}\frac{n_*}{\mu} \quad =$$

$$\frac{n_*+1}{\mu+\theta}.$$

We suppose that

$$E\big(\text{sojourn time of } n_*\text{+k -customer}\big) = \frac{n_*+k}{\mu+k\theta}.$$

$$\text{Then,}\quad E\big(\text{sojourn time of } n_*+k+1\text{-customer}\big) \quad =$$

$$\frac{1}{\mu+\big((k+1)\theta\big)} + \frac{\mu+k\theta}{\mu+\big(k+1\big)\theta}E\big(\text{sojourn time of } n_*+k\text{-customer}\big) \quad =$$

$$\frac{1}{\mu+\big(k+1\big)\theta} + \frac{\mu+k\theta}{\mu+\big(k+1\big)\theta}\cdot\frac{n_*+k}{\mu+k\theta} = \frac{n_*+k+1}{\mu+\big(k+1\big)\theta}.$$

□

Remark: In case that we interpret the n-tagged customer as the customer who finds n-customers except of him in the system then the conditional expected net benefit is

$$U\big(n|n_*\big) = \begin{cases} R - C\frac{n+1}{\mu} & \text{when } n \leq n_*-1 \\ R\frac{\mu}{\mu+\big(n+1-n_*\big)\theta} - C\frac{n+1}{\mu+\big(n+1-n_*\big)\theta} & \text{when } n \geq n_* \end{cases} \tag{1.7}$$

. By using equation $\big(1.36\big)$ we will find the expected net benefit of a customer.

Proposition 4.4: The expected net benefit of a customer that decides to join the system when the other customers follow an $\big(n_*.q\big)$ strategy and $\lambda q_* < \mu$ is given by the formula

$$\begin{aligned} U\big(n_*, q\big) &= \pi_0 R\frac{1-k_1^{n_*}}{1-k_1} - \frac{\pi_0 C}{\mu\big(1-k_1\big)^2}\Big(n_* k_1^{n_*+1} - \big(1+n_*\big)k_1^n * +1\Big) \\ &+ R\mu\Gamma\big(k_2\big)k_1^{n_*}\pi_0\Psi_{2,2}\left[\ \big(1,1\big),\big(\theta,\mu+\theta\big)\big)\big(1,k_2\big),\big(\theta,\mu+\theta+1\big)\ ;k_3\right] \\ &- \Gamma\big(k_2\big)k_1^{n_*}\pi_0 C\left(\Psi_{3,3}\left[\begin{matrix}\big(1,1\big),\big(1,1\big),\big(\theta,\mu+\theta\big)\\ \big(1,0\big),\big(1,k_2\big),\big(\theta,\mu+\theta+1\big)\end{matrix}\ ;k_3\right]\right. \\ &+ \left.\big(n_*+1\big)\Psi_{2,2}\left[\begin{matrix}\big(1,1\big),\big(\theta,\mu+\theta\big)\\ \big(1,k_2\big),\big(\theta,\mu+\theta+1\big)\end{matrix}\ ;k_3\right]\right) \end{aligned} \tag{1.8}$$

Proof. :

$$
\begin{aligned}
U\big(n_*, q\big) &= \sum_{n=0}^{\infty} \pi\big(n_*, q\big) U\big(n|n_*\big) \\
&= \sum_{n=0}^{n_*-1} \pi\big(n_*, q\big) U\big(n|n_*\big) + \sum_{n=n_*}^{\infty} \pi\big(n_*, q\big) U\big(n|n_*\big) \\
&= \sum_{n=0}^{n_*-1} k_1^n \pi_0 \big(R - c\frac{n+1}{\mu}\big) \\
&+ \sum_{n=n_*}^{\infty} \frac{\big(k_3\big)^{n-n_*} \Gamma\big(k_2\big)}{\Gamma\big(k_2 + n - n_*\big)} k_1^{n_*} \pi_0 \Big(R \frac{\mu}{\mu + \big(n+1-n_*\big)\theta} - C \frac{n+1}{\mu + \big(n+1-n_*\big)\theta}\Big) \\
&= \sum_{n=0}^{n_*-1} k_1^n \pi_0 R - \sum_{n=0}^{n_*-1} k_1^n \pi_0 c \frac{n+1}{\mu} \\
&+ \sum_{n=n_*}^{\infty} \frac{\big(k_3\big)^{n-n_*} \Gamma\big(k_2\big)}{\Gamma\big(k_2 + n - n_*\big)} k_1^{n_*} \pi_0 R \frac{\mu}{\mu + \big(n+1-n_*\big)\theta} \\
&- \sum_{n=n*}^{\infty} \frac{\big(k_3\big)^{n-n_*} \Gamma\big(k_2\big)}{\Gamma\big(k_2 + n - n_*\big)} k_1^{n_*} \pi_0 c \frac{n+1}{\mu + \big(n+1-n_*\big)\theta}.
\end{aligned}
\tag{1.9}
$$

We compute the four series separately. $\big(1\big)$:

$$
\sum_{n=0}^{n_*-1} k_1^n \pi_0 R = \pi_0 R \frac{1 - k_1^{n_*}}{1 - k_1}
$$

(2):

$$
\begin{aligned}
\sum_{n=0}^{n_*-1} k_1^n \pi_0 C \cdot \frac{n+1}{\mu} &= \sum_{n=0}^{n_*-1} k_1^n \pi_0 C \cdot \frac{n}{\mu} + \sum_{n=0}^{n_*-1} k_1^n \pi_0 \cdot \frac{C}{\mu} \\
&= \frac{\pi_0 C k_1}{\mu} \cdot \frac{\left(d\frac{1-k_1^{n_*}}{1-k_1} \right)}{dk_1} \frac{\pi_0 C}{\mu} \cdot \frac{1-k_1^{n_*}}{1-k_1} \\
&= \frac{\pi_0 C k_1}{\mu} - n_* k_1^{n_*-1}\big(1-k_1\big) + 1 - k_1^{n_*} 1 - k_1^2 + \frac{\pi_0 C}{\mu} \cdot \frac{1-k_1^{n_*}}{1-k_1} \\
&= \frac{\pi_0 C \Big(-n_* k_1^{n_*} + n_* k_1^{n_*+1} + k_1 - k_1^{n_*+1} + \big(1-k_1\big)\big(1-k_1^{n_*}\big) \Big)}{\mu \big(1-k_1\big)^2} \\
&= \frac{\pi_0 C}{\mu \big(1-k_1\big)^2} \Big(n_* k_1^{n_*+1} - \big(1+n_*\big) k_1^{n_*} + 1 \Big).
\end{aligned}
$$

(3):

$$\sum_{n=n_*}^{\infty} \frac{\big(k_3\big)^{n-n_*}\Gamma\big(k_2\big)k_1^{n_*}\pi_0}{\Gamma\big(k_2+n-n_*\big)} \frac{R\mu}{\mu+\big(n+1-n_*\big)\theta} =$$

$$= \quad R\mu\Gamma\big(k_2\big)k_1^{n_*}\pi_0 \sum_{k=0}^{\infty} \frac{\big(k_3\big)^k}{\Gamma\big(k+k_2\big)+\big(\theta k+\mu+\theta\big)}$$

$$= \quad R\mu\Gamma\big(k_2\big)k_1^{n_*}\pi_0 \sum_{k=0}^{\infty} \frac{\Gamma\big(k+1\big)\Gamma\big(\theta k+\mu+\theta\big)}{\Gamma\big(k+k_2\big)\Gamma\big(\theta k+\mu+\theta+1\big)} \frac{\big(k_3\big)^k}{k!}$$

$$= \quad R\mu\Gamma\big(k_2\big)k_1^{n_*}\pi_0 \Psi_{2,2}\left[\begin{array}{l} \big(1,1\big)\big(\theta,\mu+\theta\big) \\ \big(1,k_2\big),\big(\theta,\mu+\theta+1\big) \end{array} ;k_3\right] \tag{1.10}$$

(4):

$$\sum_{n=n_*}^{\infty} \frac{\big(k_3\big)^{n-n_*}\Gamma\big(k_2\big)k_1^{n_*}\pi_0}{\Gamma\big(k_2+n-n_*\big)} \frac{C\big(n+1\big)}{\mu+\big(n+1-n_*\big)\theta}$$

$$= \Big(\Gamma\big(k_2\big)k_1^{n_*}\pi_0 c\Big) \sum_{k=0}^{\infty} \frac{k_3^k}{\Gamma\big(k+k_2\big)} \frac{\big(k+n_*+1\big)}{\mu+\big(k+1\big)\theta}$$

$$= \Big(\Gamma\big(k_2\big)k_1^{n_*}\pi_0 c\Big)\Bigg(\sum_{k=0}^{\infty} \frac{k_3^k k}{\Gamma\big(k+k_2\big)\big(\mu+\big(k+1\big)\theta\big)} + \big(n_*+1\big) \sum_{k=0}^{\infty} \frac{k_3^k}{\Gamma\big(k+k_2\big)}\Bigg)$$

$$= \Big(\Gamma\big(k_2\big)k_1^{n_*}\pi_0 c\Big)\Bigg(\sum_{k=0}^{\infty} \frac{\Gamma\big(k+1\big)\Gamma\big(k+1\big)\Gamma\big(\theta k+\mu+\theta\big)}{\Gamma\big(k\big)\Gamma\big(k+k_2\big)\Gamma\big(\theta k+\mu+\theta+1\big)} \frac{k_3^k}{k!}$$

$$+ \quad \big(n_*+1\big) \sum_{k=0}^{\infty} \frac{\Gamma\big(k+1\big)\Gamma\big(\theta k+\mu+\theta\big)}{\Gamma\big(k+k_2\big)\Gamma\big(\theta k+\mu+\theta+1\big)} \frac{k_3^k}{k!}\Bigg)$$

$$= \Gamma\big(k_2\big)k_1^{n_*}\pi_0 c\Bigg(\Psi_{3,3}\Big[\ \big(1,1\big),\big(1,1\big),\big(\theta,\mu+\theta\big)\big(1,0\big),\big(1,k_2\big),\big(\theta,\mu+\theta+1\big)\ ;k_3\Big]$$

$$+ \quad \big(n_*+1\big)\Psi_{2,2}\left[\begin{array}{l} \big(1,1\big),\big(\theta,\mu+\theta\big) \\ \big(1,k_2\big),\big(\theta,\mu+\theta+1\big) \end{array} ;k_3\right]\Bigg). \tag{1.11}$$

By (1), (2), (3), (4) the result immediately yields.

□

In the following Lemma we will prove a monotonicity property for the expected net benefit.

Lemma 4.1 The conditional expected net benefit $U\big(n|n_*\big)$ is a strictly decreasing function of n, for any fixed $n_* \leq \frac{R\mu}{C} + \frac{\mu}{\theta}$.

Proof. : When $n \leq n_*$, $U\big(n|n_*\big) = R - \frac{Cn}{\mu}$, which is a decreasing function of n. When $n \geq n_* + 1$ then

$$U\big(n|n_*\big) = \Big(R\mu - Cn\Big)\frac{1}{\mu + \big(n - n_*\big)\theta}.$$

$$\begin{aligned}
\frac{\partial U\big(n|n_*\big)}{\partial n} &= -\frac{R\mu\theta}{\Big(\mu + \big(n - n_*\big)\theta\Big)^2} - \frac{\Big(C\big(\mu + \big(n - n_*\big)\theta\big) - Cn\theta\Big)}{\Big(\mu + \big(n - n_*\big)\theta\Big)^2} \\
&= \frac{-R\mu\theta - c\mu + Cn_*\theta}{\Big(\mu + \big(n - n_*\big)\theta\Big)^2}.
\end{aligned}$$

So, we get that

$$\begin{aligned}
\frac{\partial U\big(n|n_*\big)}{\partial n} &\leq 0 \iff \\
Cn_*\theta &\leq R\mu\theta + C\mu \iff \\
n_* &\leq \frac{R\mu}{C} + \frac{\mu}{\theta}.
\end{aligned} \tag{1.12}$$

If $n \geq \big(n_* + 1\big)$, then $U\big(n|n_*\big)$ is a decreasing function if and only if

$$n_* \leq \big(\frac{R\mu}{C} + \frac{\mu}{\theta}\big).$$

It remains to show that the function remains decreasing at the turning point from the first branch to the second, i.e., that

$$
\begin{aligned}
\Big(R - \frac{Cn_*}{\mu}\Big) &\geq \Big(\frac{R\mu - C\big(n_* + 1\big)}{\mu + \theta}\Big) \iff \\[2em]
\Big(\big(\mu + \theta\big)\big(R - \frac{Cn_*}{\mu}\big)\Big) &\geq \Big(R\mu - C\big(n_* + 1\big)\Big) \iff \\[2em]
\Big(\mu R - Cn_* + \theta R - \frac{Cn_*\theta}{\mu}\Big) &\geq \Big(R\mu - \big(Cn_* + C\big)\Big) \iff \\[2em]
\Big(\frac{Cn_*\theta}{\mu}\Big) &\leq \Big(\theta R - C\Big) \iff \\[2em]
n_* &\leq \Big(\frac{\mu R}{C} + \frac{\theta}{\mu}\Big).
\end{aligned}
$$

□

Proposition 4.5: The continuous time Markov chain $\{N(t)\}$ of the number of customers in the queueing system, when the customers follow an (n_*, q_*) strategy is

(i) Stochastically increasing in the effective arrival rate λq_*

(ii) Stochastically decreasing in the service rate μ

(iii) Stochastically decreasing in the announcement rate θ

(iv) Stochastically increasing in the reneging threshold n_*

when the others parameters are kept fixed in each case (i)-(iv).

Proof. : We will use a sample path approach.

$\left(\mathrm{i}\right)$: We consider two stochastic processes

$$\left\{N\left(t\right), \lambda q_{*}, \mu, \theta, n_{*}\right\}$$

and

$$\left\{N\left(t\right)^{'}, \left(\lambda q_{*}\right)^{'}, \mu', \theta', n_{*}'\right\}.$$

We assume that $\mu' = \mu$, $\theta' = \theta$, $\left(n_{*}\right)' = n_{*}$ and $\left(\lambda q_{*}\right)' \geq \lambda q_{*}$. We construct a coupling $\left(\left\{N_1\left(t\right)\right\}, \left\{N_2\left(t\right)\right\}\right)$ of the corresponding processes $\left\{N\left(t\right)\right\}$, $\left\{N\left(t\right)\right\}^{'}$ that record the number of customers in the system as following. The service completions are generated by the same Poisson process with rate μ'. Similarly, the announcement times are identical in the two systems and are generated by the same Poisson process with rate θ'. The arrivals at both systems are generated by a Poisson process with rate $\left(\lambda q_{*}\right)'$. However, in the first system the arrivals occur with probability $\frac{\left(\lambda q_{*}\right)'}{\lambda q_{*}}$ in order to ensure that the arrivals in the stochastic process $\left\{N\left(t\right)\right\}$ are generated by a Poisson process with rate λq_{*}, whereas in the second system the arrival occurs almost surely. We will prove that $N \preccurlyeq N^{'}$. Equivalently we will show that if there is a time t_0 so that $N\left(t_0\right) = n, N'\left(t_0\right) = n'$ with $n \leq n'$, then $N\left(t\right) \leq N'\left(t\right) \quad \forall t \geq t_o$. More concretely, we analyze all the possible cases:

$1)$ If $N_2\left(t\right) = 0$ and $N_1\left(t\right) = 0$ then at the epoch $t^{'}$ of a Poisson arrival at rate λ we get that

$$N_2\left(t'\right) = 1$$

and

$$N_1\left(t'\right) = \begin{cases} 0 & \text{with probability } 1 - \frac{\lambda q_{*}}{\left(\lambda q_{*}\right)'} \\ 1 & \text{with probability } \frac{\lambda q_{*}}{\left(\lambda q_{*}\right)'} \end{cases} \tag{1.13}$$

.

$2)$ We suppose that $N_2\left(t\right) = n', N_1\left(t\right) = n$, $n \leq n'$ and $n', n \geq 0$. Then we have the following sub-options:
$\mathrm{a})$If $n, n' \leq n_{*}$ then at the epoch of a service completion the processes $N_1\left(t\right), N_2\left(t\right)$ move to n-1 and $n^{'}$-1, respectively. At the epoch of the position announcement the number of customers remains the same in both of the systems. What is more,

at the epoch a Poisson arrival $t^{''}$ then, $N_2\big(t''\big) = n' + 1$ and

$$N_1\big(t''\big) = \begin{cases} n & \text{with probability } 1 - \frac{\lambda q_*}{\left(\lambda q_*\right)'} \\ n+1 & \text{with probability } \frac{\lambda q_*}{\left(\lambda q_*\right)'} \end{cases} \tag{1.14}$$

$\big(\text{b}\big)$ If $n, n' \geq n_*$ then at the epoch of arrival or service completion the results are the same as in the case (a). On the other hand, when an announcement occurs at time $t^{'}$ then $N_1\big(t'\big) = n - 1$ and $N_2\big(t'\big) = n' - 1$. Therefore, in each case $N_2 \geq N_1$.

c$\big)$ If $n \leq n_*$ and $n' \geq n_*$ then at the epoch of service completion or arrival, the results are the same as in the two cases above. At the time of the position announcement, in the stochastic process $\big\{N_1\big(t\big)\big\}$ the number of customers will remain the same, but in the $\big\{N_2\big(t\big)\big\}$ will be $n^{'} - 1$. Therefor, $N_1 \leq N_2$. In case that $N_1\big(t\big) = n_*$ and $N_2\big(t\big) = n_* + 1$ then in the next announcement N_1, N_2 will be equal. Furthermore, their position will be equal to n_* so the result yields immediately from the case 2-a.

$\big(\text{iii}\big)$ We consider two stochastic processes

$$\big\{N\big(t\big), \lambda q_*, \mu, \theta, n_*\big\}$$

and

$$\big\{N^{'}\big(t\big), \big(\lambda q_*\big), \mu', \theta', n_*'\big\}.$$

We assume that $\mu' = \mu$, $\big(n_*\big)' = n_*$, $\big(\lambda q_*\big)' = \lambda q_*$. We consider that the announcement rate in the stochastic process N' is higher than N, i.e. $\theta \leq \theta'$ and we will show that $N \preccurlyeq N^{'}$. We consider a coupling $\Big(N_1\big(t\big), N_2\big(t\big)\Big)$ of the process $\Big\{N(t)\Big\}$ and $\Big\{N'(t)\Big\}$ with performance measures $(\lambda q_*)', \mu', \theta'$ and reneging threshold n_*'. In order to make the position announcement rate of the initial stochastic process be exponentially distributed at rate θ, we assume that in $N_1\big(t\big)$ the inter-arrival times between the announcement of the position are exponentially distributed random variables at rate θ' but a customer will learn his position with probability $\frac{\theta}{\theta'}$. We assume that $N_2\big(t\big) = n' \leq n = N_1\big(t\big)$ $\big(1\big)$ and we will show that the stochastic ordering in $\big(1\big)$ is true for every $t \in \mathbb{R}^+$.

We have the following cases:

(a): We consider that $N_1(t) = n' < n = N_1(t)$ and $n_* < n' < n$. We consider the exponentially distributed times $Z_1 \sim\ exp\Big((n' - n_*)\theta\Big)$, which represent the time till the first customer among the $n^{'} - n_*$ learns his position and $Z_2 \sim exp\Big((n - n')\theta\Big)$ the time till the first among the $n - n^{'}$ customers learns that his position is over $n^{'}$. We have two extra sub-options. If $Z_1 < Z_2$ then N_2 will be reduced by one almost surely and N_1 will be reduced by 1 with probability $\frac{\theta}{\theta'}$ and with probability $1 - \frac{\theta}{\theta'}$ will remain the same. Therefore $N_2 < N_1$.
If $Z_2 < Z_1$ then N_2 will not change, as there is no announcement in this stochastic process. Whereas, N_1 will be reduced by one with probability $\frac{\theta}{\theta'}$ or it will remain the same with probability 1- $\frac{\theta}{\theta'}$. Based on the previous description we observe that in the stochastic process $N_2(t)$ the announcements will happen only if $Z_1 < Z_2$, while in the stochastic process $N_1(t)$ the position announcements will be in the minimum time between Z_1 and Z_2. Therefore, the announcement rate in the stochastic process is $(n - n_*)\theta'$. The assumption that the customer finally learns his position with probability $\frac{\theta}{\theta'}$ ensures that the announcement rate in the initial stochastic process $\{N(t)\}$ is $(n - n_*)\theta$.
(b):
If $N_2(t) = n' \leq n_* < n = N_1(t)$, then at the epoch of the position announcement the stochastic process N_2 will not change, due to the assumption that $n' \leq n_*$. On the other hand, N_1 will be reduced by 1 with probability $\frac{\theta}{\theta'}$ or it will not change with probability 1-$\frac{\theta}{\theta'}$. In any case we observe that $N_2(t) \leq N_1(t)$. In case of an arrival or a service completion, the desired stochastic ordering is true due to the hypothesis that the parameters λ and μ are equal.

$\square$

Remark: An equivalent interpretation of the service rate μ, is the number of service completions per time unit given that the server is continuously busy.

Corollary 4.1: The rate of service completions is a decreasing function of the announcement rate θ, for any strategy (n_*, q) of the customers and for any other fixed parameters λq, μ, R and C.

Proof. : The mean rater of service completion is $\mu\big(1 - \pi_0(n_*, q)\big)$. Using the proposition above we get that the number of customers in the system is a decreasing function of θ. Therefore, the stationary probability of the system being empty is an increasing function of θ. Therefore, the probability $1-\pi_0\big(n_*,q\big)$ is a decreasing function of θ, and the result follows. $\square$

Lemma 4.2: The expected net benefit $U\big(n_*,q\big)$ is a strictly decreasing function of q_*, for any fixed $n_* \leq \frac{R\mu}{C}+\frac{\mu}{\theta}$ and for any fixed other parameters λq, μ, θ, R and C.

Proof. : As the number of customers in the system is an increasing function of q, then, by supposing that $q_1 < q_2$ we get $N_{q_1} < N_{q_2}\big(*\big)$. Additionally, $U\big(n|n_*\big)$ is a decreasing function of n when $n_* \leq \frac{R\mu}{C}+\frac{\mu}{\theta}$. Thus, by the inequality $\big(*\big)$ yields that,

$$U\Big(N_{q_1}|n_*\Big) > U\Big(N_{q_2}|n_*\Big).\big(**\big)$$

We observe that

$$U\big(n|n_*\big) = \begin{cases} R - C\frac{n}{\mu} & \text{when } n \leq n_*, n \geq 1 \\ \\ R\frac{\mu}{\mu+\big(n-n_*\big)\theta} - C\frac{n}{\mu+\big(n-n_*\big)\theta} & \text{when } n \geq n_*+1 \end{cases} \tag{1.15}$$

depends on q only via the number of the customers in the system. Therefore if N_q is the number of the customers in the system and n_* is fixed then

$$U\big(n|n_*\big) = U\big(N_q\big).$$

But $E\Big(U\big(N_q\big)\Big) = \sum_{n=0}^{\infty}\pi\big(N_q=n\big)U\big(n|n_*\big)\ \big(***\big)$. By the equality $\big(***\big)$ we get that

$$E\big(U\big(N_q\big)\big) = U\big(n_*,q\big).$$

The inequality $\big(**\big)$ immediately yields that

$$E\Big(U\Big(N_{q_1}|n_*\Big)\Big) \ > \ E\Big(U\Big(N_{q_2}|n_*\Big)\Big) \iff$$

$$U\Big(n_*,q_1\Big) \ > \ U\Big(n_*,q_2\Big).$$

$\square$

We can now establish the existence and uniqueness of the equilibrium of the customers.

Theorem (4.1)**.** *: An equilibrium strategy always exists and is unique. We have the following cases:*

Case I: $U\big(n_e, 0\big) \leq 0$. *Then,* $\big(n_e, 0\big)$ *is the unique equilibrium strategy.*

Case II: $U\big(n_e, 1\big) < 0 < U\big(n_e, 0\big)$.*Then the equation* $U\big(n_e, q\big) = 0$ *has exactly one root* q_e *with respect to* q *in* $\big(0, 1\big)$*, and* $\big(n_e, q_e\big)$ *is the unique equilibrium strategy.*

Case III: $U\big(n_e, 1\big) \geq 0$. *Then* $\big(n_e, 1\big)$ *is the unique equilibrium strategy.*

Proof. At the epoch of his position announcement a n-tagged customer will stay in the system in case that his net benefit is not negative. Therefore, he will leave the system if $R - C\frac{n}{\mu} < 0 \iff n > \frac{R\mu}{C} \iff n > \lfloor \frac{R\mu}{C} \rfloor$. As a result, the best strategy of a tagged customer against any strategy of the others is to use the reneging threshold $n_e = \lfloor \frac{R\mu}{C} \rfloor$. As the equilibrium strategies are best responses against themselves, so necessarily any equilibrium strategy $\big(n_*, q\big)$ should have $n_* = n_e$.

Case I: We have already seen that the function $U\big(n_e, q\big)$ is a decreasing function of q. Therefore, if $\max\limits_{0 \leq q \leq 1} U\big(n_e, q\big) \leq 0 \iff U\big(n_e, 0\big) \leq 0$ an individual's profit who decides to enter the system will be negative even if he finds the system empty. Thus, if $U(n_e, 0) < 0$ then the best response is to leave the system, so $q = 0$ the only best response against itself. Additionally if $U\big(n_e, 0\big) = 0$ then the n-tagged customer is indifferent about joining the queue or balk thus, every $q \in [0, 1]$ is a best response. But, because all the other customers follow the strategy $\big(n_e, 0\big)$, only $\big(n_e, 0\big)$ is best response against itself, consequently it is the unique equilibrium strategy.

Case III: If $\min\limits_{0 \leq q \leq 1} U\big(n_e, q\big) \geq 0 \iff U\big(n_e, 1\big) \geq 0$, then a customer that decides to enter the system will leave positive net benefit even if all the customers in front of him entered the system. Therefore, the net benefit of an individual that will enter the system will be positive for any $q \in [0,1]$. But the only best response against itself is the strategy $\big(n_e, 1\big)$.

Case II: If $\min\limits_{0\leq q\leq 1} U(n_e, q) < 0 < \max\limits_{0\leq q\leq 1} U(n_e, q) \iff U\big(n_e, 1\big) < 0 < U\big(n_e, 0\big)$. Then $\big(n_e, 1\big)$ is not best response since, in case all the customers in the system follow $\big(n_e, 1\big)$ strategy, then the net benefit of a customer who decides to enter the system w.h.p $\big($q=1$\big)$ is negative. On the other hand, under the strategy $\big(n_e, 0\big)$ the benefit of an incoming customer is positive, therefore his join probability will be positive $\big(q > 0\big)$, so $\big(n_e, 0\big)$ is not an equilibrium strategy. Since the function $U\big(n_e, q\big)$ is continuous and decreasing in q then by Bolzano's theorem a unique root q_e exists such that $U\big(n_e, q_e\big) = 0$. As $U\big(n_e, q\big)$ is a decreasing function of q then, $U\big(n_e, q\big) > 0 \ \ \forall q \in [0, q_e)$, so the best response of a tagged customer is $\big(n_e, 1\big)$. If $q \in \big(q_e, \infty\big)$ then $U\big(n_e, q\big) < 0$, therefore the customer will not enter the system and $\big(n_e, 0\big)$ is the best strategy. As $U\big(n_e, q_e\big) = 0$ therefore his best response is $\big(n_e, q\big)$ for every $q \in [0, 1]$. But, the only strategy that is best response against itself is $\big(n_e, q_e\big)$. $\square$

In the following section we will explore the effect of the announcement rate θ on the equilibrium join probability q_e and the equilibrium social welfare. As a first step, we are studying the effect of θ on the expected net benefit of a tagged customer $U(n_e, q_*)$ when the other parameters of the system are kept fixed and the customers follow a fixed strategy (n_e, q_*). The result is established formally in the following Lemma.

Lemma 4.3 The expected net benefit $U(n_e, q_*)$ is a strictly increasing function of θ, for other fixed parameters λ, μ, R, C, $q_* \neq 0$ and $n_e = \lfloor \frac{\mu R}{C} \rfloor$.

Proof

We use a coupling approach, so we consider an arriving tagged customer and his evolution in two systems $\{N(t), \lambda, \mu, \theta_1, n_*, q_*\}$ and $\{N^{'}(t), \lambda^{'}, \mu, \theta_2, n^{'}_*, q^{'}_*\}$. We assume that $\mu = \mu^{'}$, $\lambda q_* = \lambda q^{'}_*$ and $n_* = n^{'}_*$ while $\theta_1 < \theta_2$. We construct a coupling of the corresponding processes $\{N(t)\}$ and $\{N^{'}(t)\}$ that record the number of customers in the two system as following: We consider a two dimensional stochastic process $\Big\{N_1(t) \times N_2(t), \lambda q_*, \mu, \theta_2, n_*\Big\}$, i.e. the service completions are identical in the two systems and are generated by the same Poisson process with rate μ. Similarly, the arrivals are identical in the two systems and are generated by the same Poisson process with rate λq_*. On the other hand, for system 1, we assume that an announcement occurs by a Poisson process with rate $\{\theta_2\}$. Whereas, for system 2, we assume that an announcement occurs at an event of the Poisson process at rate $\{\theta_2\}$ with probability $\frac{\theta_1}{\theta_2}$. This ensures that the arrivals at system $\{N^1(t), \lambda, \mu, \theta_2, n_*, q_*\}$ are identical with the process $\{N(t), \lambda, \mu, \theta_1, n_*, q_*\}$. In particular, we generate the numbers of customers in the system at the instant arrival of the tagged customer in the two systems, which we denote as $N^1(t_0)$and $N^2(t_0)$ for system 1 and 2, respectively, so that $N^2(t_0) \leq N^1(t_0)$. This is possible as we have proved that the number of customers in the queue is stochastically decreasing function in the announcement rate θ. Moreover, we couple the service times in the two systems after the arrival of the tagged customer. Therefore, the tagged customer's potential service completion time (if he does not renege) will be longer in system 1 than in system 2. Consider, now, the first announcement epochs after the tagged customer's arrival. According to the assumptions of the model, the interarrival time between a tagged customer's arrival and his position announcement after his arrival, is exponential(θ_2) distributed in both systems, but we assume that in system 1 the announcement happens with probability $\frac{\theta_1}{\theta_2}$. We denote by t_2 and t_1 the first announcement epoch after the tagged customer's arrival in systems 2 and 1, respectively. Because of the coupling construction, we have that $t_2 \leq t_1$. If the tagged customer has been served in the system 2 by t_2, then it is clear that he is better off in system 2. If he is still in the system 2 at time t_2, then we consider two cases: $t_1 = t_2$ (case 1) and $t_1 > t_2$ (case 2).

Case 1: $t_1 = t_2$ with probability $\frac{\theta_1}{\theta_2}$, i.e. the first announcement occurs at the same time in both systems, let $Y^1(t)$ and $Y^2(t)$ be the number of customers in front of the tagged customer at time t, for systems 1 and 2. Because of the coupling of the service times in the two systems and the fact $N^2(0) \leq N^1(0)$, we have that $Y^2(t) \leq Y^1(t)$. If $n_e < Y^2(t)$, the tagged customer will renege from the first system almost surely while he will renege from system the second system if the announcement occurs (with probability $\frac{\theta_1}{\theta_2}$) thus the the cost of the customer at system 1 will be at least as low as the cost in system 2. If $Y^2(t) \leq n_e < Y^1(t)$ the tagged customer will renege from system 1 but not from system 2, so his expected net benefit will be better in system 2 (by the definition of n_e). If $Y^2(t) \leq Y^1(t) \leq n_e$, then the tagged customer will stay in both systems but he is better off in system 2 since he will suffer less waiting cost in it. If $t_1 > t_2$, i.e. the first announcement after his arrival at system 2 occurs before the one in system 1, then for all subcases similar arguments show that the tagged customer will again be better off in system 2. Therefore, we conclude that the expected net benefit of the tagged customer in system 2 is greater that the corresponding expected net benefit in system 1.

We are now in position to prove that the equilibrium join probability is an increasing function of the announcement rate.

Theorem (4.2)**.** *: In the $M/M/1$ queue with independent delayed observations, the unique equilibrium join probability $q_e(\theta)$ is a non-decreasing function of the announcement rate θ. In particular when $\lambda < \mu$, we have the following cases as R varies in $(0, \infty)$.*

Case I: $R \leq \frac{C}{\mu}$. *Then $q_e(\theta) = 0$ for $\theta \in (0, \infty)$.*

Case II: $\frac{C}{\mu} < R < \frac{C}{\mu - \lambda}$. *Then,*

$$\lim_{\theta \to 0}(q_e(\theta)) = \frac{1}{\lambda}\left(\mu - \frac{C}{R}\right) \in (0, 1)$$

and $q_e(\theta)$ is increasing for $\theta \in (0, \infty)$. Moreover, there exists a $\theta_o > 0$ such that $q_e(\theta) < 1$ for $\theta \in (0, \theta_o)$ while $q_e(\theta) = 1$ for $\theta \in (0, \infty)$.

Case III: $R \geq \frac{c}{\mu - \lambda}$. Then $q_e(\theta) = 1$, for $\theta \in (0, \infty)$.

When $\lambda \geq \mu$ we still have case I as above, but case II corresponds to $R > \frac{c}{\mu}$, *while case III does not exist.*

Proof. Case I: If the join probability in the system is $q_* = 0$ then the utility function of tagged arriving customer will be

$$U(n_e, 0; \theta) = \begin{cases} U(0, 0; \theta) & \text{when } n_e = 0 \\ U(n_e, 0; \theta) & \text{when } n_e > 0 \end{cases} \tag{1.16}$$

If $n_e = 0$ then as $q_* = 0$ the customer will find the system empty but he will complete his service only if his service completion happens before the announcement of his position. Therefore, $U(0, 0; \theta) = \frac{R\mu}{\mu+\theta} - \frac{C}{\mu+\theta} \geq 0$ because $R \geq \frac{C}{\mu}$. In case that $n_e > 0$ then $R\mu \geq C$. On the other hand, according to hypothesis $R\mu \leq C$, so we conclude that $n_e = 1 \iff R\mu = C$. Therefore, $U(n_e, 0; \theta) = U(1, 0; \theta) = R - \frac{C}{\mu} = 0$. In any case $U(n_e, 0; \theta) \leq 0 \quad \forall n_e \in \mathbb{N}$. Therefore, in case that $U(n_e, 0, \theta) < 0$ then $q_e = 0$ is the unique equilibrium strategy. If $U(n_e, 0; \theta) = 0$ then every $q \in [0, 1]$ is best response for the n-tagged customer but only $q_e = 0$ is an equilibrium strategy. Thus, for every $\theta \in (0, \infty)$, $q_e = 0$ is the unique equilibrium strategy.

Case III: Let's suppose that the customers enter the system with probability $q_* = 1$. We know that as $\theta \to 0$ the $M/M/1$ with independent delayed observations agrees with the model of Edelson and Hildebrand. Therefore,

$$\lim_{\theta \to 0} U(n_e, q_*; \theta) = R - \frac{C}{\mu - \lambda q_*}.$$

Thus, for $q_* = 1$

$$\lim_{\theta \to 0} U(n_e, 1; \theta) = R - \frac{C}{\mu - \lambda} \geq 0.$$

By the previous Lemma we know that $U(n_e, 1; \theta)$ is an increasing function of θ, so $U(n_e, 1; \theta) \geq 0 \ \forall \theta > 0$. Therefore, $q_e = 1$ is a symmetric best response for the n-tagged customer and hence the unique equilibrium strategy.

Case II: As $R - \frac{C}{\mu} > 0$ then $n_e > 0$. If the join probability is $q_* = 0$, then the n-tagged customer will definitely receive his reward so $U(n_e, 0; \theta) = R - \frac{c}{\mu} > 0$ $\forall \theta \in (0, \infty)$. Thus, his best response is $q_e(\theta) = 1$ which is not an equilibrium strategy. Regarding the behavior of U for $q = 1$ we have that

$$\lim_{\theta \to 0} U(n_e, 1; \theta) = R - \frac{C}{\mu - \lambda} < 0.$$

On the other hand

$$\lim_{\theta \to \infty} U(n_e, 1; \theta)$$

is the utility function of a customer that joins the system in Naor's model, therefore it is positive. We know that $U(n_e, 1; \theta)$ is a continuous function of θ. Thus, there exists $\theta_0 \in (0, \infty)$ such that $U(n_e, 1, \theta_0) = 0$. As $U(n_e, 1, \theta)$ is an increasing function of θ, thus we get $U(n_e, 1, \theta) \geq 0 \;\; \forall \theta \geq \theta_0$ and $U(n_e, 1, \theta) < 0 \; \forall \theta < \theta_0$. For $\theta > \theta_0$, $q_e(\theta) = 1$ is a equilibrium strategy for the n-tagged customer. On the other hand for $\theta \in (0, \theta_o)$, $U(n_e, 1, \theta) < 0$. Also, $U(n_e, 0; \theta) > 0 \forall \theta \in (0, \theta_0)$. As $U(n_e, q, \theta)$ is continuous in θ there exists a unique root q_o such that $U(n_e, q_o; \theta) = 0 \forall \theta \in (0, \theta_o)$. As U is decreasing in q and increasing in θ then we get that $q_o(\theta)$ is increasing in θ. Therefore, in this case the function $q_e(\theta)$ is increasing in θ, starting from $\lim_{\theta \to 0} q_e(\theta) = \frac{1}{\lambda}\mu - \frac{C}{R} \in (0, 1)$ (unobservable model) and reaching $q_e(\theta) = 1$ for $\theta = \theta_0$ and $\forall q > q_e$ is 1.

$\square$

Proposition 4.6 When the system is centrally controlled and a strategy (n_*, q) is somehow imposed, the social mean net benefit per time unit is given by

$$S(n_*, q_*) = \lambda q_* U(n_*, q)$$

where

$$
\begin{aligned}
U\big(n_*, q\big) &= \pi_0 R\frac{1-k_1^{n_*}}{1-k_1} - \frac{\pi_0 C}{\mu\big(1-k_1\big)^2}\Big(n_* k_1^{n_*+1} - \big(1+n_*\big)k_1^n * +1\Big) \\
&+ R\mu\Gamma\big(k_2\big)k_1^{n_*}\pi_0\Psi_{2,2}\left[\ \big(1,1\big),\big(\theta,\mu+\theta\big)\big)\big(1,k_2\big),\big(\theta,\mu+\theta+1\big)\ ;k_3\right] \\
&- \Gamma\big(k_2\big)k_1^{n_*}\pi_0 C\left(\Psi_{3,3}\left[\begin{array}{l}\big(1,1\big),\big(1,1\big),\big(\theta,\mu+\theta\big)\\ \big(1,0\big),\big(1,k_2\big),\big(\theta,\mu+\theta+1\big)\end{array};k_3\right]\right. \\
&+ \left.\big(n_*+1\big)\Psi_{2,2}\left[\begin{array}{l}\big(1,1\big),\big(\theta,\mu+\theta\big)\\ \big(1,k_2\big),\big(\theta,\mu+\theta+1\big)\end{array};k_3\right]\right)
\end{aligned}
\tag{1.17}
$$

Appendix Mittag-Leffler function

The one-parametric Mittag-Leffler function

Definition. The one-parametric Mittag-Leffler function is defined by the power series

$$E_a\big(z\big) = \sum_{k=0}^{\infty} \frac{z^k}{\Gamma\big(ak+1\big)}, \qquad a \in \mathbb{C}$$

Theorem. $\Big($***Cauchy-Hadamard theorem for one complex variable z***$\Big)$
Consider the formal power series in one complex variable z of the form

$$f(z) = \sum_{n=0}^{\infty} c_n (z-a)^n, a, c_n \in \mathbb{C}.$$

Then the radius of f at the point a is given by

$$R = \frac{1}{\limsup\limits_{n\to\infty}(|c_n|^{\frac{1}{n}})} \quad or,$$

$$R = \frac{1}{\limsup\limits_{n\to\infty}\left(\frac{|C_n|}{C_{n+1}}\right)}\Bigg)$$

Theorem. $\Big($***Cauchy-Hadamard theorem for several complex variables***$\Big)$
Let a be a multi-index with $|a| = a_1 + \cdots + a_n$*, then f(x) converges with radius of convergence* ρ *(which is also multi-index) if and only if* $\lim_{|a|\to\infty}\Big(|C_a|\rho^a\Big)^{\frac{1}{|a|}} = 1$ *to the multidimensional power series*

$$\sum_{a_1\geq 0, a_2\geq 0, \ldots, a_n\geq 0} C_{a_1,\ldots,a_n}\big(z_1 - a_1\big)^{a_1} \ldots \big(z_n - a_n\big)^{a_n}$$

By using the asymptotic expansion of the book "Transcendental function", page 73 1.18(5) we get that

$$\frac{\Gamma(z+\alpha)}{\Gamma(z+\beta)} = z^{\alpha-\beta}\Big[1+\frac{z^{-1}(\alpha-\beta)(\alpha+\beta-1)}{2}+O(z^{-2})\Big]$$

, thus, according to Cauchy - Hadamard formula for the radius of convergence, where $C_k = \frac{1}{\Gamma(\alpha k+1)}$

$$\limsup_{n\to\infty}\Big(\frac{|C_n|}{|C_{n+1}|}\Big) = \limsup_{n\to\infty}(a_n^{a+1-1})[1+\frac{(a+1-1)(a+1-1-1)}{2an}+O(\frac{1}{a^2n^2})] \iff$$

$$\limsup_{n\to\infty}\Big(\frac{|C_n|}{|C_n+1|}\Big) = \limsup_{n\to\infty}((an)^a)[1+\frac{(a-1)}{2n}+O(\frac{1}{a^2n^2})](*).$$

In case that $Re(a) > 0$ then

$$(*) = \limsup_{n\to\infty}((an)^a)[1+\frac{(a-1)}{2n}+O(\frac{1}{a^2n^2})] = \infty.$$

In case that $Re(a) < 0$ then

$$(*) = \limsup_{n\to\infty}((an)^a)[1+\frac{(a-1)}{2n}+O(\frac{1}{a^2n^2})] = 0.$$

If $Re(a) = 0$ it can be proved that the radius of convergence is, $e^{\frac{\pi|Ima|}{2}}$.

Definition. In complex analysis, an entire function, also called an integral function, is a complex-valued function that is holomorphic at all finite points over the whole complex plane.

Therefore, in case that $Re(a) > 0$ the Mittag-Leffler function is an entire function.

For specific values of a, the Mittag-Leffler function is equal to some special functions. For example:

- $E_0(\pm z) = \sum_{k=0}^{\infty}(\pm 1)^k z^k$.
- $E_1(\pm z) = \sum_{k=0}^{\infty} \frac{\pm 1)^k z^k}{\Gamma(k+1)} = e^{\pm z}$.
- $E_2(-z^2) = \sum_{k=0}^{\infty} \frac{(-1)^k z^{2k}}{\Gamma(2k+1)} = cosz$
- $E_2(z^2) = \sum k = 0^{\infty} \frac{z^{2k}}{\Gamma(2k+1)} = coshz$

Proposition A more general formula for the function with half-integer parameter is valid,

$$E_{p/2}(z) = F_{0,p-1}\left(1/p, 2/p, ..., \frac{p-1}{p}; \frac{z^2}{p^p}\right) + \frac{2^{\frac{p+1}{2}} z}{p!\sqrt{\pi}} F_{1,2p-1}\left(\frac{p+2}{2p}, \frac{p+3}{2p}, ..., \frac{3p}{2p}; \frac{z^2}{p^p}\right),$$

where

$$F_{p,q}(z) = F_{p,q}(a_1, a_2, a_3, ..., a_p; b_1, b_2, ..., b_1, b_2, ..., b_q; z) = \sum_{k=0}^{\infty} \frac{(a_1)_k (a_2)_k \ldots (a_p)_k}{(b_1)_k (b_2)_k \ldots (b_q)_k} \frac{z^k}{k!}$$

the (p,q)-hypergeometric function.

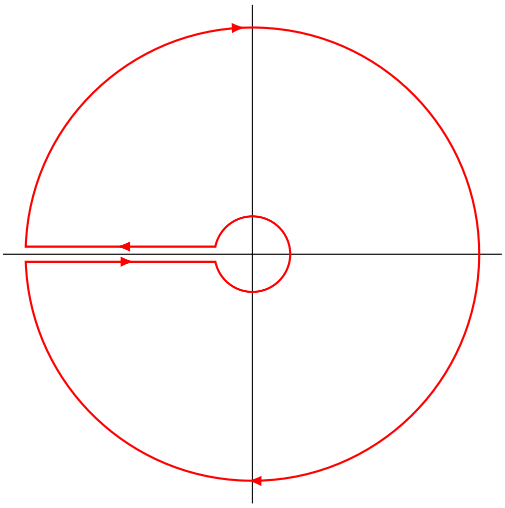
Proposition

The Mittag-Leffler function satisfies the following differential relations :

$$\left(\frac{d}{dz}\right)^p E_p\left(z^p\right) = E_p\left(z^p\right)$$

$\frac{\partial^p}{\partial z^p} E_{p/q}(z^{p/q}) = E_{p/q}(z^{p/q}) + \sum_{k=1} q - 1 \frac{z^{\frac{-kp}{q}}}{\Gamma(1-\frac{kp}{q})}, \quad q = 2, 3, ...$

Many important properties of the Mittag-Leffler function follow from its integral representations. By denoting $\gamma\big(\epsilon, \alpha\big)$, $\big(\epsilon > 0, 0 < \alpha \geq \pi\big)$ a contour oriented by non-decreasing argz consisting of the following parts: the ray $argz = -a$, $|z| \leq \epsilon$, the $arc - a \leq argz \leq a, |z| = \epsilon$, and the ray $argz = a$, $|z| \geq \epsilon$. If $0 < a < \pi$, then the contour $\gamma(\epsilon; a)$ divides the complex z-plane into two unbounded parts, namely $G^{(-)}(\epsilon, \alpha)$ to the left of $\gamma(\epsilon, \alpha)$ by orientation, and $G^{(+)}(\epsilon, \alpha)$ to the right of it. In case that, $a = \pi$, then the contour consists of the circle $|z| = \epsilon$ and the twice passable ray $-\infty < z \leq -\epsilon$. In both cases the contour $\gamma(\epsilon; a)$ is called the Hankel path.



Proposition Let $0 < a < 2$ and $\frac{\pi a}{2} < \beta \leq min\{\pi, \pi a\}$.

Then the Mittag-Leffler function can be represented in the form

$$E_a(u) = \frac{1}{2\pi a i} \int_{\gamma(\epsilon,\beta)} \frac{e^{z^{\frac{1}{a}}}}{z-u} du \quad , \quad u \in G^-(\epsilon, \beta)$$

$$E_a(u) = \frac{1}{a} e^{z^{\frac{1}{a}}} + \frac{1}{2\pi a i} \int_{\gamma(\epsilon,\beta)} \frac{e^{z^{\frac{1}{a}}}}{z-u} du \quad , \quad u \in G^+(\epsilon, \beta)$$

Based on the integral representation of the function there is a relation between the Mittag-Leffler function and the Fox-Wright function, known as the Euler transform of the Mittag- Leffler function. Let $\alpha, \quad \rho, \quad \sigma \quad \in \mathbb{C}$, $\gamma > 0$ and $Re\alpha > 0$, $Re\sigma > 0$, then the following representation holds:

$$\int_0^1 x^{\rho-1}(1-x)^{\sigma-1} E_a(tx^{\gamma})\, dx = \Gamma(\sigma)\Psi_{2,2}\left[\begin{array}{l} (\rho,\gamma),(1,\alpha) \\ (1,\alpha),(\sigma+\rho,\gamma) \end{array} ;t \right],$$

where $\Psi_{2,2}$ is a special case of the Fox-Wright function $\Psi_{p,q}$:

$\Psi_{2,2}\left[\begin{array}{l} (\rho,\gamma),(1,\alpha) \\ (1,\alpha),(\sigma+\rho,\gamma) \end{array} ;x \right] = \sum_{k=0}^{\infty} \frac{\Gamma(\rho+\gamma k)\Gamma(1+k)}{\Gamma(1+\alpha k)\Gamma(\sigma+\rho+\gamma k)} \frac{x^k}{k!}$

The two-parametric Mittag-Leffler function

Definition. The two-parametric Mittag-Leffler function is defined by

$$E_{\alpha,\beta}(z) = \sum_{k=0}^{\infty} \frac{z^k}{\Gamma(\alpha k + \beta)} \quad (\alpha, \beta \in \mathbb{C})$$

and is an entire function for $\mathrm{Re}(\alpha) > 0$

Using the above definition we obtain a number of formulas relating the two-parametric Mittag-Leffler function to elementary functions:

$$E_{1,2}(z) = \frac{e^z - 1}{z}$$

$$E_{2,1}(z) = cosh\sqrt{z}$$

$$E_{2,2}(z) = \frac{sinh\sqrt{z}}{\sqrt{z}}.$$

The two parametric Mittag-Leffler function is an entire function and satisfies the the following differential formulas, The Mittag-Leffler function satisfy the follownig defferential relations :

$$\left(\frac{d}{dz}\right)^p [z^{\beta-1} E_{p,\beta}\big(\lambda z^p] = z^{\beta-p-1} E_{p,\beta-p}\big(\lambda z^p\big), (n \in \mathbb{N}, \lambda \in \mathbb{C}) and$$

$$\left(\frac{d}{dz}\right)^p [z^{p-\beta} E_{p,\beta}(\frac{\lambda}{z^p}] = \frac{(-1)^p \lambda}{z^{n+\beta}} E_{p,\beta}(\frac{\lambda}{z^p}), (z \neq 0; n \in \mathbb{N}; \lambda \in \mathbb{C})$$

The integral representation of the two-parametric Mittag-Leffler function is:

$$E_{\alpha,\beta}(u) = \frac{1}{2\pi i\alpha} \int_{\gamma(\epsilon,\delta)} \frac{e^{z^{\frac{1}{\alpha}}} z^{\frac{(1-\beta)}{\alpha}}}{z-u} dz \quad , \quad u \in G^{-}(\epsilon,\delta)$$

$$E_{\alpha,\beta}(u) = \frac{1}{\alpha} z^{\frac{(1-\beta)}{\alpha}} e^{z^{\frac{1}{\alpha}}} + \frac{1}{2\pi i\alpha} \int_{\gamma(\epsilon,\delta)} \frac{e^{z^{\frac{1}{\alpha}}} z^{\frac{(1-\beta)}{\alpha}}}{z-u} dz \quad , \quad u \in G^{+}(\epsilon,\delta)$$

under the condition $0 < \alpha < 2$, $\frac{\pi\alpha}{2} < \delta < \min(\pi, \pi\alpha)$. The contour $\gamma(\epsilon,\delta)$ consists of two rays $S_{-\delta}(argz = -\delta, |z| \geq \epsilon)$ and $S_{\delta}(argz = \delta, |z| \geq \epsilon)$ and a circular arc $C_{\delta}(0,\epsilon)(|z| = \epsilon, -\delta \leq argz \leq \delta)$. On its left side there is a region $G^{-}(\epsilon,\delta)$ and on its right side a region $G^{+}(\epsilon,\delta)$.

References

- Adan Ivo and Jacques Resing (2015), *Queueing Systems*, Department of Mathematics and Computing Science-Eindhoven University of Technology.

- Burnetas Apostolos, Antonis Economou, George Vasiliadis, (2017), 'Strategic customer behavio *Queueing systems.*

- Cooper Robert, *Introduction to QUEUEING THEORY, second edition*, North Holland, New York.

- Edelson N.M and K. Hildebrand (1975), 'Congestion tolls for Poisson queueing processes', *Econometrica.*

- Gorenflo Rudolf, Anatoly A. Kilbas, Francesco Mainardi, Sergei V. Rogosin, *Mittag-Leffler Functions, Related Topics and Applications*, Springer.

- Hassin Refael, Moshe Haviv, (2003), *TO QUEUE OR NOT TO QUEUE, Equilibrium Behavior in Queueing Systems*, Kluwer Academic Publisher, Boston.

- Hassin Refael (1986), 'Consumer information in markets with random products quality: The c
Econometrica.

- Hassin Refael, Ricky Roet-Green, (2013), 'The armchair decision: depart or stay home', 2013, working paper.

- Kilbas Anatoly, Srivastava Hari, Trujillo Juan, *Theory and Applications of Fractional Differential equations*, ELSEVIER B.V.

- Kulkarni Vidyadhar, *Modeling and Analysis of Stochastic Systems. Second edition*, CRC Press

- Naor P. (1969) 'The regulation of queue size by levying tolls', *econometrica.*

- Puterman Martin, (2008), *Markov Decision Processes: Discrete Stochastic Dynamic Programming*, Wiley Series in Probability and Statistics.