Università degli Studi di Padova

DIPARTIMENTO DI MATEMATICA "TULLIO LEVI-CIVITA"

Corso di Laurea Magistrale in Matematica

Equilibrium strategies for overtaking free queueing networks under partial information

Relatore:
Prof. Bernardo D'Auria

Laureando: Alberto Cesaro
Matricola: 2045233

Anno Accademico 2022/2023

21/07/2023

Contents

List of Figures

Abstract

[English] In this thesis, we are interested in finding equilibrium strategies for customers arriving at overtaking free queueing networks, only knowing partial information about the state of the system. The overtaking free condition does not allow customers to be overtaken by those behind them. We suppose that customers arrive at the system according to a Poisson process and that their service times at any queue are independent and exponentially distributed. Upon her arrival, the tagged customer is informed about the total number of customers in the system and chooses whether to join or not. Assuming that all customers follow the same strategy, the aim of the thesis is to find the equilibrium strategy that gives the maximum profit for any arriving customer. We show that such a strategy exists and is a pure or mixed threshold strategy. After analyzing the two-node tandem network and the multi-node tandem network, we focus on queueing networks with a branching structure, the so called tree networks. The work is extended with some numerical calculations, simulations and examples of non-overtaking free networks.

[Italiano] In questa tesi, siamo interessati a trovare strategie di equilibrio per clienti che arrivano a reti di code *overtaking free*, conoscendo solo informazioni parziali riguardanti lo stato del sistema. La condizione di *overtaking free* non permette ai clienti di essere superati da quelli dietro di loro. Supponiamo che i clienti arrivino al sistema secondo un processo di Poisson e che i loro tempi di servizio ad ogni coda siano indipendenti e distribuiti esponenzialmente. Al suo arrivo, il cliente in questione viene informato del numero totale di clienti nel sistema e sceglie se unirsi o meno. Assumendo che tutti i clienti seguano la stessa strategia, l'obiettivo della tesi è trovare la strategia di equilibrio che dia il massimo profitto per ogni cliente in arrivo. Mostriamo che una tale strategia esiste ed è una strategia di tipo soglia pura o mista. Dopo aver analizzato la rete tandem a due nodi e la rete tandem a più nodi, ci concentriamo sulle reti di code con una struttura ramificata, le cosiddette reti ad albero. Il lavoro viene ampliato con alcuni calcoli numerici, simulazioni ed esempi di reti non *overtaking free*.

Introduction

Queueing theory is a field of mathematics widely studied to understand how the flow of people or objects occurs in a queueing network. In general we do not like to wait and, before standing in line, each of us would like to know whether the time spent waiting is worth the service we will receive. So, in this thesis, we are interested in finding equilibrium strategies for customers arriving at overtaking free queueing networks and receiving partial information. Such a strategy says whether it is convenient for a customer to join or not the system based on her expected sojourn time at any queue. The overtaking free condition does not allow customers to be overtaken by those behind them, which means that the sojourn time of a customer is not affected by arrivals after her.

Agner Krarup Erlang and Tore Olaus Engset are the real founders of the queueing theory. In particular, in 1909, Erlang started to work on the waiting times in a telephone exchange and he identified that the number of telephone conversations satisfied a Poisson distribution as well as the telephone waiting time was exponential distributed [8]. Indeed, we also assume that customers arrive at the system according to a Poisson process and that their service times are identically and exponentially distributed. Within these assumptions, we study different overtaking free models in order to find the optimal strategy to be adopted by arriving customers, who are informed only about the total number of customers in the system.

Chapter 1 is a collection of preliminary results. It contains definitions, models and examples, which will be useful throughout the thesis. We analyze the Jackson networks, which are the general structures of a queueing system, and look at the stationary distribution of customers. Some results about stochastic orders, Poisson processes and equilibrium strategies are discussed. Then, we focus on the Naor's model [15], which consists of one single server queue, at which customers arrive, observe its length and decide whether to join or not. We also describe the Little's law, that relates the average sojourn time and the average number of customers to the arrival rate at the system. Finally, we recall some notions about graphs and trees.

Chapter 2 studies the two-node tandem network, which consists of two queues in series [7]. It means that customers joins the system at the first queue and, after completing the service, they enter the second queue. When a customer arrives, she is informed about the total number of customers in the system and has to choose whether to join or not. For this reason the aim of this chapter is to find the equilibrium strategy to be adopted by all customers in order to maximize their profit function, which depends on the sojourn time in each queue. Some numerical computations allow us to see how the equilibrium strategy changes when some parameters vary. Finally, we simulate the flow of customers in the two-node tandem network.

Chapter 3 is a generalization of the previous one. We consider a multi-node tandem network, that is a system with M queues arranged in tandem [12]. We analyze the expected sojourn time at each queue in order to find the equilibrium strategy that says whether joining the system is convenient or not for an arriving customer. The assumption is that all customers use the same strategy. In this way, following the equilibrium strategy is the best choice for any customer arriving at the system. Then, we compute the values of the equilibrium threshold as the service rates change.

Chapter 4 deals with tree networks. Tree networks are particular queueing systems where customers, after receiving the service at a certain queue, can move on to different queues, but at a certain queue customers arrive from only one queue. It means that the queues are arranged as an out-tree. As in the previous chapters, in this one we compute the equilibrium strategy that gives the better payoff for any arriving customer. Then, we compare this model with a tandem network and analyze an example of tree network.

Chapter 5 concerns the overtaking-free condition, which is satisfied by previously mentioned models. First, we characterize the overtaking queueing networks with exponential service times, under a *first-in-first-out* discipline. Then we analyze a model which is not overtaking free under the FIFO discipline, but it becomes so if the service discipline is changed. However, we also study an example of non overtaking free network. Finally, we change the distribution of the service times. We consider a model with one only queue and Erlang service times, and look for an equilibrium strategy.

Chapter 1

Preliminaries

Abstract

In this chapter we give some preliminary results, that will be useful later, about: Jackson networks [5], stochastic orderings [5], Poisson processes [3, 21, 14], equilibrium and threshold strategies [9], the Naor's model [9, 15], the Little's law [18, 4, 13], and something about graphs and trees [2, 6].

1.1 Jackson networks

In this section we give an overview of Jackson networks, which are very useful structures to describe queueing networks. The main reference is [5].

A Jackson network consists of J nodes, each with one or several servers. The service rate at each node i can be both node-dependent and state-dependent. Specifically, whenever there are x_i jobs at node i, the service rate is $\mu_i(x_i)$, where $\mu_i(\cdot)$ is a function $\mathbb{Z}_+ \longmapsto \mathbb{R}_+$ with $\mu_i(0) = 0$ and $\mu_i(x_i) > 0$ $\forall x_i > 0$. Jobs travel among the nodes following a routing matrix $P := (p_{ij})_{i,j=1}^J$, where p_{ij} is the probability that a job leaving node i will go to node j. At each node all jobs are served on a *first-in-first-out* basis, that is there is no priority in serving the jobs.

According to different specifications of the routing matrix, there are three different variations: the *open*, *closed* and *semiopen* network.

Open network

In an open network jobs arrive from outside following a Poisson process with rate $\lambda > 0$. Each arrival is independently routed to node j with probability $p_{0j} \geq 0$, with $\sum_{j=1}^J p_{0j} = 1$. It means that each node j has an independent external Poisson stream of arrivals with rate λp_{0j}. Upon service completion at node i, a job may go to another node j with probability p_{ij} or leave the network with probability $1 - \sum_{j=1}^J p_{ij}$.

Let λ_i be the overall arrival rate to node i, then we have the following traffic equation

$$\lambda_i = \lambda p_{0i} + \sum_{j=1}^{J} \lambda_j p_{ji}, \quad i = 1, \ldots, J, \tag{1.1}$$

which can be written in matrix notation in the following way

$$a = b + P^T a,$$

where $a := (\lambda_i)_{i=1}^J$, $b := (\lambda p_{0i})_{i=1}^J$ and $P := (p_{ji})_{i,j=1}^J$. So, assuming that $I - P^T$ is invertible, we have that

$$a = (I - P^T)^{-1} b.$$

Let $Q_i(t)$ denote the number of jobs at node i at time t. Then $\{(Q_i(t))_{i=1}^J,\ t \geq 0\}$ is a continuous-time Markov chain. Set $Q = (Q_i)_{i=1}^J$ and $x = (x_i)_{i=1}^J$. The Markov chain is governed by the following transition rates, denoted by $q(\cdot,\cdot)$: for $x \in \mathbb{Z}_+^J$

$$\begin{aligned}
q(x, x + e_i) &= \lambda p_{0i}, \\
q(x, x - e_i) &= \mu_i(x_i) p_{i0}, \\
q(x, x - e_i + e_j) &= \mu_i(x_i) p_{ij} \\
-q(x, x) &= \sum_{i=1}^{J} [\lambda p_{0i} + \mu_i(x_i)(1 - p_{ii})]
\end{aligned}$$

and 0 otherwise, where e_i is the i-th J-dimensional unit vector.

Let $\pi(x)$ denote the equilibrium (steady-state) distribution, that is $\pi(x) = \mathbb{P}(Q = x)$. Then it satisfies the relation

$$\pi^t A = 0, \tag{1.2}$$

where A is the *rate matrix*, whose entries $q(x, y)$ correspond to the transition rates between each pair of states x and y specified above. Generally, relation (1.2) is known as *global balance equations*, which equates the "probability flow" out of each state x with the flow into the same state:

$$\pi(x) \sum_{y \neq x} q(x, y) = \sum_{y \neq x} \pi(y) q(y, x), \quad \forall x \in \mathbb{Z}_+^J.$$

In this case the relation to be satisfied is

$$\begin{aligned}
\pi(x) \sum_{i=1}^{J} [\lambda p_{0i} + \mu_i(x_i)(1 - p_{ii})] &= \\
= \sum_{i=1}^{J} [\pi(x - e_i) \lambda p_{0i} + \pi(x + e_i) \mu_i(x_i + 1) p_{i0}] &+ \\
+ \sum_{i=1}^{J} \sum_{j \neq i} \pi(x + e_i - e_j) \mu_i(x_i + 1) p_{ij}, &\quad \forall x \in \mathbb{Z}_+^J.
\end{aligned} \tag{1.3}$$

The main result below, Theorem 1.1, relates the stationary distribution of Q to a vector of independent random variables $Y = (Y_1, \dots, Y_J)$, where each Y_i has a probability mass function as follows:

$$\mathbb{P}(Y_i = n) = \mathbb{P}(Y_i = 0) \cdot \frac{\lambda_i^n}{\mu_i(1) \cdots \mu_i(n)}. \tag{1.4}$$

We assume that

$$\sum_{n=1}^{\infty} \frac{\lambda_i^n}{\mu_i(1) \cdots \mu_i(n)} < \infty, \tag{1.5}$$

so that $\mathbb{P}(Y_i = 0)$ is well-defined, namely,

$$\mathbb{P}(Y_i = 0) = \left[1 + \sum_{n=1}^{\infty} \frac{\lambda_i^n}{\mu_i(1) \cdots \mu_i(n)}\right]^{-1}.$$

The following theorem says how to compute the stationary distribution and it can be proved substituting the expression of $\pi(x)$ in the system of balance equations (1.3). For the proof, see [5].

Theorem 1.1 (Stationary distribution for open networks)**.** Provided that condition (1.5) holds for all $i = 1, \dots, J$, the stationary distribution of the number of jobs in an open Jackson network is

$$\pi(x) = \prod_{i=1}^{J} \mathbb{P}(Y_i = x_i),$$

for $x \in \mathbb{Z}_+^J$, where Y_i follows the distribution in (1.4).

Closed network

In a closed network the total number of jobs in the network is maintained at a constant level, say N. Once a job completes all of its processing requirements and leaves the network, a new job is immediately released into the network. Conceptually, this type of operation can also be viewed as having a fixed number of jobs circulating in the network, with no job ever leaving the network and no external job entering the network, and in this sense the network is "closed". So, in a closed model, the routing matrix $P := (p_{ij})_{i,j=1}^J$ is *stochastic*, that is the row sums are all equal to one. In other words

$$p_{i0} = p_{0j} = 0 \quad \forall i, j = 1, \dots, J,$$

using the notation of the open model, because no job enters into the system and no job leaves.

Let v_i be the arrival rate to node i, i.e. the counter part of λ_i in the open model. It follows that $(v_i)_{i=1}^J$ is the solution to the following traffic equation

$$v_i = \sum_{j=1}^{J} v_j p_{ji}, \quad i = 1, \dots, J.$$

The solution of the above system is unique only up to a multiplier constant and we need another equation, for example imposing $\sum_{j=1}^{J} v_j = 1$ or $v_i = 1$ for some node i.

The equilibrium balance equations are similar to those of the open network, except that here $\lambda = 0$ and $p_{i0} = p_{0i} = 0$ for all $i = 1, \ldots, J$. Hence, we have

$$\pi(x) \sum_{i=1}^{J} \mu_i(x_i)(1 - p_{ii}) = \sum_{i=1}^{J} \sum_{j \neq i} \pi(x + e_i - e_j)\mu_i(x_i + 1)p_{ij},$$

for all $x \in \mathbb{Z}_+^J$ such that $|x| = N$, where $|x| = x_1 + \cdots + x_J$.

The following theorem gives an expression to compute the stationary distribution. For the proof, we refer to [5].

Theorem 1.2 (Stationary distribution for closed networks)**.** The closed Jackson network, with a total of N jobs, has the following stationary distribution

$$\pi(x) = \prod_{i=1}^{J} \frac{\mathbb{P}(Y_i = x_i)}{\mathbb{P}(|Y| = N)},$$

for $x \in \mathbb{Z}_+^J$ such that $|x| = N$, where Y_i follows the distribution in (1.4) with $x_i \leq N$ and λ_i replaced by v_i.

Following the scheme of the open model, we don't have to require the condition (1.5) and the denominator in $\pi(x)$ comes from the normalizing condition, since

$$\mathbb{P}(|Y| = N) = \sum_{|x|=N} \prod_{i=1}^{J} \mathbb{P}(Y_i = x_i),$$

while in the open network the normalizing condition is

$$\sum_{x \in \mathbb{Z}_+^J} \prod_{i=1}^{J} \mathbb{P}(Y_i = x_i) = 1.$$

Semiopen network

A semiopen network follows the description of the open model with the exception that the total number of jobs in the network is limited to a maximum of K jobs at any time. When this limit is reached, external arrivals are blocked and lost. The arrival process will be resumed when a job leaves the network, bringing the total number of jobs in the network down to $K - 1$.

It turns out that the semiopen model can be reduced to a closed network with a constant of K jobs and $J + 1$ nodes. The additional node, indexed as node 0, again represents the external world, with routing from and to node 0 following the probabilities p_{0j} and p_{i0}, just as in the open network. Hence, in this case, the routing matrix of this closed network with $J+1$ nodes is $P := (p_{ij})_{i,j=0}^J$. In addition, node 0 is given a service function, with "service" rate $\mu_0(n) = \lambda$ for all $n \geq 1$ and $\mu_0(0) = 0$. In this way, node 0 effectively generates the arrival process of the original open network, blocking the external arrivals when the network is full.

Let λ_i be the arrival rate to node i, then it satisfies the following traffic equation

$$\lambda_i = \lambda p_{0i} + \sum_{j=1}^{J} \lambda_j p_{ij}, \quad i = 1, \ldots, J,$$

and, dividing both sides by λ and letting $v_0 = 1$, $v_i = \lambda_i/\lambda$ for $i = 1, ..., J$, we obtain that $(v_i)_{i=0}^J$ satisfies

$$v_i = \sum_{j=0}^{J} v_j p_{ji}, \quad i = 0, 1, \ldots, J,$$

which is the traffic equation of the closed model for a network with $J+1$ nodes, with the additional equation $v_0 = 1$.

Using this observation, we can state the following theorem, that gives an expression to compute the stationary distribution for a semiopen network [5].

Theorem 1.3 (Stationary distribution for semiopen networks)**.** The semiopen Jackson network, with an overall buffer capacity of K, has the following stationary distribution

$$\pi(x) = \prod_{i=1}^{J} \frac{\mathbb{P}(Y_i = x_i)}{\mathbb{P}(|Y| \leq K)},$$

for $x \in \mathbb{Z}_+^J$ such that $|x| \leq K$, where Y_i follows the distribution in (1.4) with $x_i \leq K$.

1.2 Stochastic orderings

In this section we give some definitions about stochastic orderings and their relation.

Definition 1.4 (Usual stochastic ordering)**.** Let X and Y be two discrete random variables with the same support set $\mathcal{N}$. We say that X dominates Y in *usual stochastic ordering*, denoted by $X \geq_{st} Y$, if and only if

$$\mathbb{P}(X \geq n) \geq \mathbb{P}(Y \geq n), \quad \forall n \in \mathcal{N}.$$

Remark 1.5 (Characterization of usual stochastic ordering). It can be verified that $X \geq_{st} Y$ if and only if, for all non decreasing function $h(\cdot)$, $\mathbb{E}[h(X)] \geq \mathbb{E}[h(Y)]$.

Definition 1.6 (Likelihood ratio ordering)**.** Let X and Y be two discrete random variables with the same support set $\mathcal{N}$. We say that X dominates Y in *likelihood ratio ordering*, denoted by $X \geq_{\ell r} Y$, if and only if

$$\mathbb{P}(X = n)\mathbb{P}(Y = n-1) \geq \mathbb{P}(X = n-1)\mathbb{P}(Y = n), \quad \forall n \in \mathcal{N}.$$

The following lemma says that the usual stochastic ordering is, actually, *weaker* than the likelihood ratio ordering.

Lemma 1.7 (Usual and likelihood ratio orderings)**.** Let X and Y be two discrete random variables with the same support set $\mathcal{N}$. Then

$$X \geq_{\ell r} Y \Rightarrow X \geq_{st} Y.$$

For the proof, see [5].

1.3 Poisson processes

An homogeneous Poisson process is a countable random set of points of the real line with some properties. In many applications, a point of an homogeneous Poisson process is the time of occurrence of some events, for example the arrival times of customers to a queue. In the sequel, we simply use "Poisson process" to refer to an homogeneous Poisson process. So, in these chapters, a Poisson process represents the customer arrival process at a queueing system. The main reference of this section is [3].

Definition 1.8 (Random point process)**.** A *random point process* is a sequence $(\tau_n)_{n\geq 0}$ of nonnegative random variables such that, almost surely, $0 = \tau_0 < \tau_1 < \tau_2 < \dots$ and $\lim_{n\to\infty} \tau_n = +\infty$.

The sequence $(\delta_n)_{n\geq 1}$ defined by

$$\delta_n = \tau_n - \tau_{n-1}, \quad n \geq 1,$$

is called the *interarrival* sequence and it represents the time that passes between the arrival of a customer and another.

The family of random variables $N = (N(t))_{t\geq 0}$, with

$$N(t) = \sum_{n=1}^{+\infty} \mathbb{1}_{\{\tau_n \in (0,t]\}}, \quad t \geq 0,$$

is call the *counting process* of the point process $(\tau_n)_{n\geq 0}$ and it counts the number of customers arriving at a queueing system in the time interval $(0, t]$. Since the sequence of arrival customers can be recovered from N, the latter is also called "point process".

Definition 1.9 (Poisson process)**.** A counting process N is called a *Poisson process* with intensity $\lambda > 0$ if

(i) for all times $0 \leq t_1 \leq t_2 \leq \dots \leq t_k$, the random variables $N(t_{i+1}) - N(t_i)$, for $i = 1, \dots, k-1$, are independent;

(ii) for any $a < b$, with $a, b \in \mathbb{R}_+$, $N(b) - N(a)$ is a Poisson random variable with mean $\lambda(b - a)$.

Since $N(t) \sim \text{Poi}(\lambda t)$, then $\mathbb{E}[N(t)] = \lambda t$, that is λt is the average number of customers arrived at a queueing system in a time period equal to t.

Theorem 1.10 (Interarrivals are i.i.d. and exponential)**.** The interarrival sequence $(\delta_n)_{n\geq 1}$ of a Poisson process with intensity $\lambda > 0$ is i.i.d. and exponentially distributed with parameter λ.

So, $\delta_n \sim \text{Exp}(\lambda)$ and $\mathbb{E}[\delta_n] = \lambda^{-1}$, that is λ^{-1} is the average time that passes between an arrival of a customer and another.

Sometimes customers arrive at a queueing system following different independent Poisson processes. The following theorem explains what happens in this situation.

Theorem 1.11 (Sum of independent Poisson processes)**.** Let $\{N_i\}_{i\geq 1}$ be a family of independent Poisson processes with respective positive intensities $\{\xi_i\}_{i\geq 1}$. Then, two distinct Poisson processes of this family have no points in common and, if $\lambda := \sum_{i=1}^{+\infty} \xi_i < +\infty$,

$$N(t) := \sum_{i=1}^{+\infty} N_i(t)$$

defines the point process of a Poisson process with intensity λ.

For the proofs of Theorem 1.10 and Theorem 1.11, see [3]. The following theorem is concerned with the PASTA property, which stands for *Poisson Arrivals See Time Averages*. The PASTA property refers to the expected state of a queueing system as seen by an arrival from a Poisson process. A proof of a simplified version of Theorem 1.12 is in [14]. For more details, see [21].

Theorem 1.12 (PASTA property)**.** An arrival from a Poisson process observes the system as if it was arriving at a random moment in time.

Therefore, the expected value of any parameter of the queue at the instant of a Poisson arrival is simply the long-run average value of that parameter, since the average state of the system is the steady state.

Remark 1.13 (Why should we use a Poisson process for arriving customers?)*.* There is a theoretical justification for the reason why the Poisson process is a good approximation of the customer arrival process in many real-world systems.

We consider a huge universe of people that arrive at a queueing network and suppose that each person flips a coin, which decides whether that person becomes a customer, that might join the system, or not. It means that each person behaves as a Bernoulli trial. So, in this model, the number of people that become potential customers depends on how many of these independent Bernoulli trials result in success.

The probability of getting k successes out of N independent Bernoulli trials with success parameter p is given by the binomial distribution. If N is large and p is small – as they are in our model – the binomial distribution can be approximated by the Poisson distribution. Therefore, the probability of getting k arrivals in a period is approximately Poisson distributed, which implies that the arrival process is also nearly Poisson.

1.4 Equilibrium and threshold strategies

In this section, we give some notions about strategies in non-cooperative games, looking at the definition of equilibrium strategy. Then, we focus on a class of strategies, the so-called *threshold strategies*. The main reference is [9].

Definition 1.14 (Pure and mixed strategy)**.** 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 player i is an action from A_i. A *mixed strategy* for player i corresponds to a probability function which prescribes a randomized rule for selecting an action from A_i.

Denote by S_i the set of strategies available to player i. A *strategy profile* $s = (s_1, \ldots, s_n)$ assigns a strategy $s_i \in S_i$ to each player $i \in N$.

Each player $i \in N$ is associated with a real payoff function $F_i(s)$, which is assumed to be linear in s_i. This function specifies the payoff received by player i given that the strategy profile s is adopted by the players. We denote by $s_{-i} = (s_1, \ldots, s_{i-1}, s_{i+1}, \ldots, s_n)$ a profile for the set of players $N \setminus \{i\}$.

Definition 1.15 (Best response)**.** A strategy s_i^* is said to be the *best response* for player i against the profile s_{-i} if

$$s_i^* \in \arg\max_{s_i \in S_i} F_i(s_i, s_{-i}),$$

that is

$$F_i(s_i^*, s_{-i}) \geq F_i(s_i, s_{-i}), \quad \forall s_i \in S_i.$$

Definition 1.16 (Equilibrium profile)**.** A strategy profile s^e is an *equilibrium profile* if, for every $i \in N$, s_i^e is a best response for player i against s_{-i}^e, i.e.

$$s_i^e \in \arg\max_{s_i \in S_i} F_i(s_i, s_{-i}^e), \quad \forall i \in N.$$

We will deal mostly with games with indistinguishable infinitely many players (usually customers). In this case, denote the common set of strategies and the payoff function by S and F, respectively. Let $F(a, b)$ be the payoff for a player who selects strategy a, when everyone else selects strategy b.

Definition 1.17 (Equilibrium strategy)**.** An *equilibrium strategy* is a strategy $s^e \in S$ such that

$$s^e \in \arg\max_{s \in S} F(s, s^e),$$

that is

$$F(s^e, s^e) \geq F(s, s^e) \quad \forall s \in S.$$

In other words, s^e is a equilibrium strategy if it is a best response against itself. We can also give a definition of *dominant strategy*, which is a condition stronger than being an equilibrium strategy.

Definition 1.18 (Dominant strategy)**.** A strategy $s^* \in S$ is a *dominant strategy* if it holds that

$$F(s^*, \bar{s}) \geq F(s, \bar{s}), \quad \forall s \in S, \quad \forall \bar{s} \in S.$$

Remark 1.19 (Steady-state). When evaluating an individual's expected payoff which is associated to a strategy $\bar{s}$ as a response against all others using strategy s, we assume that steady-state conditions (based on all using strategy s) have been reached, and the tagged individual assumes that the stationary distribution is the distribution over the states. Indeed, in most of the models, there is an underlying Markov process, whose transition probabilities are induced by the common strategy s selected by all.

We describe a class of strategies, known as *threshold strategies*, which is common in queueing systems. Suppose that upon arrival the customer has to choose between two actions, A_1 and A_2, after observing a non negative integer-valued variable which characterizes the state of the system. For example, the state may be the length of the queue and the actions may be to join or to balk.

Definition 1.20 (Pure threshold strategy)**.** A *pure threshold strategy* with threshold $K \in \mathbb{N}$ prescribes one of the actions, say A_1, for every state in $\{0, 1, \ldots, K-1\}$ and the other action, A_2, otherwise.

Since it is not always possible to find a pure threshold strategy, we introduce also the following extended concept.

Definition 1.21 (Threshold strategy)**.** A *threshold strategy* with threshold $x = n + p$, $n \in \mathbb{N}$, $p \in [0, 1)$, prescribes mixing between the two pure threshold strategies n and $n + 1$, so that strategy n receives the probability weight of $1 - p$ and strategy $n + 1$ receives the weight of p. The resulting behaviour is that any individual selects a given action, say A_1, when the state is $0 \leq i \leq n - 1$; she selects randomly between A_1 and A_2 when $i = n$, assigning probability p to A_1 and probability $1 - p$ to A_2; and she selects A_2 when $i > n$. If x is an integer (p=0), the strategy is *pure*, otherwise it is *mixed.*

1.5 $M/M/1$ model

We give some basic concepts about queueing theory which will be used frequently. The following definitions and results are described in [9]. The decisions made by the customers of a queueing system are to *join* or to *balk*, based on whether a customer enters the system or not. A joining customer can also decide to *renege*, when she removes herself from the queue while waiting, but such an action is not considered in our analysis. The *arrival process* corresponds to the process by which customers arrive to the system, while the *joining process* consists only of those customers who decide to join.

The service discipline used most in these chapters is *first-in-first-out* (FIFO), that is there are no priorities and customers are served based on the order of arrival at each queue. Another service discipline is *last-in-first-out* (LIFO). We distinguish a LIFO discipline *without preemption*, in which a new customer is positioned at the head of the queue, but the customer in service is allowed to complete it, from a LIFO discipline *with preemption*, in which a new arrival preempts a customer who might be in service. It is usually assumed that service, when resumed, is continued from the point it was interrupted and such discipline is also known as *preemption-resume* LIFO discipline. Where not specified, the discipline used is the FIFO one. In Chapter 5, we analyze some queueing networks with service disciplines different from the FIFO one.

In addition, we say that a queueing network is *overtaking free* if any customer, while waiting, is not overtaken by those who have arrived at the system after her. The definition of overtaking free networks is given in Section 5.1.

We give below some results for a network with one single server node with infinite buffer and service times independent and identically exponentially distributed. Let the service time distribution be exponential with rate μ, which means that the expected service time is μ^{-1}. The arrival process is a Poisson process with rate λ. The model is known as $M/M/1$ model, where M stands for *Markovian* and 1 for a single server.

We analyze the observable model and the unobservable model. In the observable model, the arriving customer is informed about the number of customers in front of her, while, in the unobservable one, the customer has no information about the state of the system. In both cases, we compute the expected sojourn time and show which is the equilibrium strategy to be adopted by the arriving customers in order to maximize their expected profit. Such a strategy says whether to join or balk.

Observable model

In the observable model, to find the equilibrium strategy, the tagged customer makes her decision by optimizing the expected profit function

$$P(k) = R - CT(k)$$

where R is the reward she gets for joining the network, C the cost she pays for each unit of sojourn time and $T(k)$ the expected sojourn time, knowing that the system contains k customers. We suppose that $R \geq C/\mu$, because otherwise, even when the system is empty, no customers would enter the system. The tagged customer decides to join if $P(k) \geq 0$ and to balk otherwise. The model with these assumptions was firstly studied by Naor and it is called *Naor's model* ([15]).

So, we suppose that all customers, when arrive at the system, are informed about the total number k of users in the network. Then the expected sojourn time is $T(k) = (k+1)/\mu$ and the expected profit function becomes

$$P(k) = R - \frac{C(k+1)}{\mu},$$

Hence, we define

$$K = \inf\{k \in \mathbb{Z}_+ : P(k) < 0\} = \left\lfloor \frac{R\mu}{C} \right\rfloor. \tag{1.6}$$

Then, the equilibrium strategy is given by the threshold strategy with threshold equal to K, that is an arriving customer joins if and only if she observes a number of customers less than K. Such a strategy is also a dominant strategy in the class of threshold strategies. This result is due to [15].

Indeed, let $F_{K_1,K_2}(k)$ be the payoff function of the tagged customer that chooses to follow the K_1-strategy, when all the others follow the K_2-strategy. In this case the payoff function is

$$F_{K_1,K_2}(k) = P(k)\mathbb{1}_{\{0\le k<K_1\}}, \quad k = 0, 1, \ldots, K_2$$

From Definition 1.18, a K^*-strategy is a dominant strategy if $F_{K^*,K_2}(k) \ge F_{K_1,K_2}(k)$ for all K_1, K_2. By the definition of the payoff function, such a strategy is implemented by the threshold K, defined in (1.6).

If $P(K-1) > 0$ (that is when $R\mu/C$ is not an integer), the K-strategy is a pure threshold strategy for this model, while, if $P(K-1) = 0$ (that is when $R\mu/C$ is an integer), the K-strategy and the $(K-1)$-strategy are pure threshold strategies and any strictly convex combination between these two strategies is a mixed threshold strategy for this model.

Remark 1.22 (Non-threshold strategies). There are no other threshold strategies for this model, but there are other non-threshold strategies. For example, we can take a strategy such that a customer joins the system if $k \ne K$ and balk if $k = K$. Indeed, since there are never more than K customers in the system, what customers decide to do when $k > K$ is irrelevant to establishing whether a strategy defines an equilibrium or not.

Unobservable model

In the unobservable model, the expected profit no longer depends on the number of customers in the queue, because the customer has no information about the state of the queue at her arrival. So the expected profit is

$$P = R - CT,$$

where T is the expected sojourn time without any information. In this case, assuming that $R > CT$, all customers enter the system and, so, the model behaves like an *open* Jackson network. Let Q^* be the stationary number of customers in the queue. For stability, we assume that $\mu > \lambda$. Then, by Theorem 1.1, the stationary distribution $\pi = (\pi(k))_{k\in\mathbb{Z}_+}$ is

$$\pi(k) = \mathbb{P}(Q^* = k) = (1-\rho)\rho^k, \quad k \in \mathbb{Z}_+,$$

where $\rho = \lambda/\mu$.

We compute the expected sojourn time

$$T = \sum_{k=0}^{+\infty} T(k)\pi(k) = \sum_{k=0}^{+\infty} \frac{k+1}{\mu}(1-\rho)\rho^k = \frac{1}{\mu(1-\rho)} = \frac{1}{\mu-\lambda}, \tag{1.7}$$

hence, the expected profit is

$$P = R - \frac{C}{\mu - \lambda}.$$

So, assuming that $R > \frac{C}{\mu-\lambda}$, the expected sojourn time under no information is $T = 1/(\mu - \lambda)$, and, as a consequence of this, the equilibrium strategy says to join the queue for any arriving customer.

Remark 1.23 (Multiple arrival processes)*.* Let us suppose that customers arrive at the system according to N independent Poisson process with intensity ξ_i, $i = 1, \dots, N$. So, we assume that there is no longer just one arrival process, but more than one, precisely N arrival processes, $N \geq 2$. We set $\lambda := \sum_i \xi_i$ and $c_i = \xi_i/\lambda$, where c_i represents the probability that an arriving customer comes from the i-th arrival process. By Theorem 1.11, we have that the superposition of the N independent Poisson processes is a Poisson process with intensity λ. So the effect of arrival customers from N arrival processes is the same of one single arrival process of rate λ. As a consequence of this, all the results previously obtained also hold for this situation.

1.6 Little's law

In a queueing network, the average sojourn time of the customers in the system can often be obtained by applying the Little's law. We introduce the Little's law with its meaning and application, then we show a short proof and an example in the unobservable $M/M/1$ model. The main references are [18, 4, 13].

Specifically, the Little's law for a service system says that the average sojourn time T of a customer and the average number of customers L in the system are related by

$$L = \bar{\lambda} T, \tag{1.8}$$

where $\bar{\lambda}$ is the arrival rate of customers to the system.

This fundamental relation is a "law of averages" or "law of large numbers" when the quantities $L, \bar{\lambda}, T$ are limits of averages. It is also a "law of expectations" when the quantities are expected values. In studying a system, one may want to use $L = \bar{\lambda} T$ to obtain one of these values from the other two. We are mainly interested in computing the quantity T, knowing $\bar{\lambda}$ and L.

We consider a general service system that processes customers. Suppose users arrive at the system at times $0 < \tau_1 < \tau_2 < \dots$, where $\tau_n \to +\infty$ a.s. We refer to the arrivals by the point process

$$N(t) = \sum_{n=1}^{+\infty} \mathbb{1}_{\{\tau_n \in (0,t]\}}, \quad t \geq 0,$$

which denotes the number of arrivals in the time interval $(0, t]$.

Let S_n denote the entire time the n-th customer is in the system, including the service time. We call S_n the *sojourn time* of the n-th customer.

The n-th customer departs from the system at time $\tau_n + S_n$ and never returns. The number of customers that arrive in the time interval $(0, t]$ and are still in the system at time t is

$$Q(t) = \sum_{n=1}^{+\infty} \mathbb{1}_{\{\tau_n \leq t < \tau_n + S_n\}}, \quad t \geq 0.$$

According to the definitions of the processes $\tau_n, S_n, Q(t)$, we set

$$\begin{aligned} L &\overset{a.s.}{=} \lim_{t\to+\infty} t^{-1} \int_0^t Q(s)ds &&\longrightarrow \quad \text{average number of customers} \\ \bar{\lambda} &\overset{a.s.}{=} \lim_{t\to+\infty} t^{-1} N(t) &&\longrightarrow \quad \text{average arrival rate} \\ T &\overset{a.s.}{=} \lim_{n\to+\infty} n^{-1} \sum_{k=1}^{n} S_k &&\longrightarrow \quad \text{average sojourn time} \end{aligned}$$

The key idea behind the Little's law $L = \bar{\lambda}T$ is that the integral of $Q(t)$ is simply another way of recording sojourn times. Specifically, if the system is empty at times 0 and t ($Q(0) = Q(t) = 0$), then the sojourn time of the customers up to time t is

$$\int_0^t Q(s)ds = \sum_{k=1}^{N(t)} S_k. \tag{1.9}$$

To better understand Eq. (1.9), we can look at Figure 1.1 (a), which describes a potential situation of customers that arrive at the system at times $\tau_1, \tau_2, \ldots$ and exit the system after a sojourn time equal to $S_1, S_2, \ldots$ In the figure, the grey area corresponding to customer i is the sojourn time S_i, so it is easy to verify that Eq. (1.9) holds.

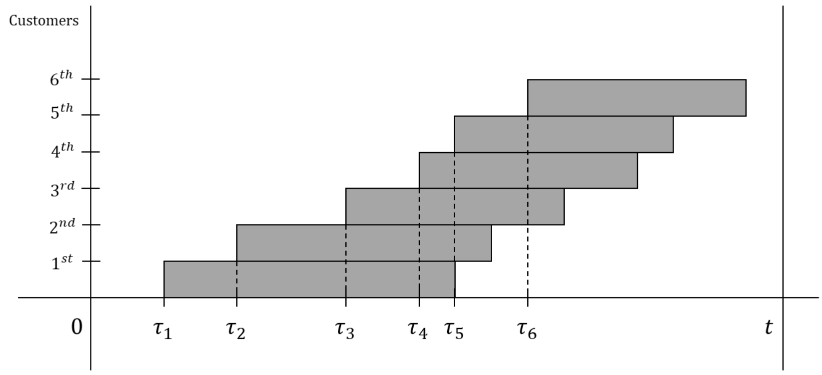
Even at nonempty times t, that is when $Q(0) = 0$ and $Q(t) > 0$, a system satisfies

$$t^{-1} \int_0^t Q(s)ds = t^{-1} \sum_{k=1}^{N(t)} S_k + o(1) \tag{1.10}$$

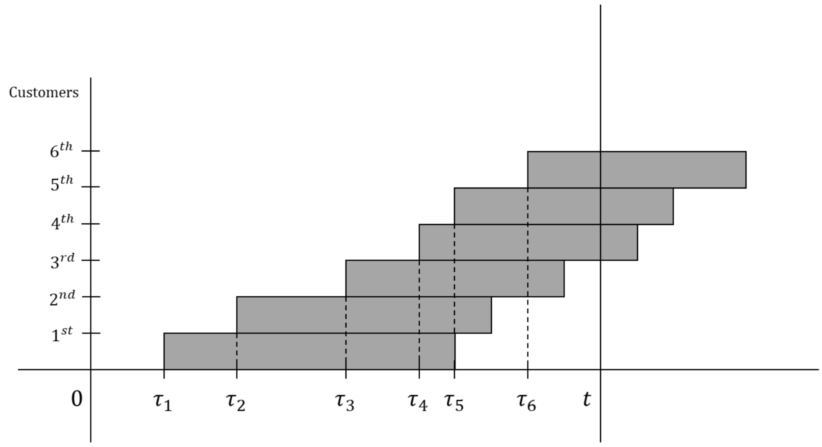
where $o(1) \to 0$ as $t \to +\infty$. For example, let us see in Figure 1.1 (b) the same situation of Figure 1.1 (a), where the time t is fixed in such a way that $Q(t) > 0$. In this case, letting $t \to +\infty$ in (1.10) yields $L = \bar{\lambda}T$.

Let us now consider the unobservable $M/M/1$ model, as described in Section 1.5, and see how the Little's law can be applied. First, we observe that all customers use the same strategy, that is to enter the system. Then, we compute the expected number of customers in the system

$$\mathbb{E}[Q^*] = \sum_{k=0}^{+\infty} k\pi(k) = \sum_{k=0}^{+\infty} k(1-\rho)\rho^k = \frac{\rho}{1-\rho} = \frac{\lambda}{\mu - \lambda},$$



(a) With empty time t



(b) With nonempty time t

Figure 1.1: Scheme of customers arriving in a system. The i-th customer arrives at time T_i and leaves the system after a sojourn time S_i.

and, using the Little's law, the expected sojourn time turns out to be equal to

$$T = \frac{\mathbb{E}[Q^*]}{\lambda} = \frac{1}{\mu - \lambda},$$

which is the same expression obtained in (1.7).

Remark 1.24 (When can we use the Little's law?). As observed in the previous example, we can apply the Little's law only if the arriving customers all use the same strategy. This assumption is due to the fact that the Little's law is a "law of averages" and, as a consequence of this, it requires all users to behave in the same way. For example, if we suppose that any arriving customer enters the system, then the Little's law is trivially applicable, but when the strategy is threshold type, we cannot compute the expected sojourn time $T(k)$, conditioned on being k customers in the system, because not all joining customers observe k customers at their arrival. So, for example, what we can compute is the expected sojourn time of customers that join if and only if they observe less than $k+1$ customers in the system at their arrival.

1.7 Graphs and trees

In queueing systems it can be often useful to work with structures such as graphs or trees to describe the flow of customers during their route. So, we give some basic concepts about graph theory ([2, 6]).

Definition 1.25 (Undirected graph)**.** An *undirected graph* G is an ordered pair $G = (V, E)$ of finite sets V and E, where E is a set of unordered pairs of elements of V.

Definition 1.26 (Directed graph)**.** A *directed graph* G is an ordered pair $G = (V, E)$ of finite sets V and E, where E is a set of ordered pairs of elements of V.

In a directed graph $G = (V, E)$, V is the set of *vertices* and E is the set of *edges*. An edge $e \in E$ is denoted with $e = (v_1, v_2)$, where $v_1, v_2 \in V$ correspond to the pair of vertices associated to e. By definition $(v_1, v_2) \neq (v_2, v_1)$. We denote with ν the number of vertices of G, $\nu = |V|$, and with ε the number of edges of G, $\varepsilon = |E|$.

For a directed graph $G = (V, E)$, we define the *indegree* $d^+(v)$ and the *outdegree* $d^-(v)$ of a vertex $v \in V$ as the number of edges entering v and the number of edges outgoing from v, that is $d^+(v) = |S^+(v)|$ and $d^-(v) = |S^-(v)|$, where $S^+(v) = \{w \in V : (w, v) \in E\}$ and $S^-(v) = \{w \in V : (v, w) \in E\}$.

In a directed graph $G = (V, E)$, we say that two vertices $v, w \in V$ are connected with a *path* if there exists a sequence of vertices $\{v = v_0, v_1, \dots, v_{n-1}, v_n = w\}$, where $v_0, \dots, v_n \in V$ and $(v_i, v_{i+1}) \in E$ for all $i = 0, \dots, n-1$. A path $\{v_0, v_1, \dots, v_{n-1}, v_n\}$ is a *cycle* if $v_0 = v_n$. A graph is *acyclic* if it contains no cycles.

Definition 1.27 (Tree)**.** An undirected graph $G = (V, E)$ is a *tree* if it is connected and acyclic.

Definition 1.28 (Out-tree)**.** Let $G = (V, E)$ be a directed graph and $r \in V$ a vertex of it. Then G is an *out-tree* if $|E| = |V| - 1$ and, for any $v \in V \setminus \{r\}$, there exists a directed path from r to v.

The vertex r is called *root* of the out-tree. The following lemma gives a characterization of out-trees.

Lemma 1.29 (Characterization of out-trees)**.** Let $G = (V, E)$ be a directed graph and $r \in V$ a vertex of it. Then G is an out-tree with root r if and only if its underlying undirected graph is a tree and it holds that $d^+(r) = 0$ and $d^+(v) = 1$ for any $v \in V \setminus \{r\}$.

Lemma 1.29 implies that there is a unique path that connects the root r to any vertex of the out-tree. A vertex v is called a *leaf* of the tree if $d^-(v) = 0$. We denote with $V_2 \subset V$ the set of leafs of a tree, and with $V_1 = V \setminus V_2$ the set of vertices with at least one outgoing edge.

Chapter 2

Two-node tandem network

Abstract

In this chapter, we show that in a two-node tandem network the arriving customers, knowing only partial information, can adopt a pure threshold strategy, which is the optimal one. This strategy requires that a customer chooses whether to join or not the queue, without knowing the complete state of the system, but only the total number of people in the system. The threshold strategy is obtained by maximizing a given profit function, whose costs are proportional to the sojourn time in each queue. Some numerical computations allow us to see the behaviour of the equilibrium threshold when some parameters vary. The main reference is [7].

2.1 The model

The model consists of a tandem network with two single server nodes with infinite buffers and service times independent and exponentially distributed, where customers are served according to a FIFO discipline. We index the nodes by l, $l = 1, 2$.

We denote by μ_l, the service rate at node l. By denoting with g_l the generic service time at node l, we have that

$$g_l \sim \text{Exp}(\mu_l) \quad \text{with} \quad g_1 \perp g_2,$$

and it follows that $\mathbb{E}[g_l] = 1/\mu_l$, that is on average μ_l customers are served per unit of time.

We suppose that customers arrive to the system according to a Poisson process with rate λ. As seen in Section 1.3, it means that, on average, a customer will arrive in λ^{-1} units of time and, after a time period t, λt customers will have arrived. The model is graphically represented in the queueing network of Figure 2.1.

The arriving customers receive partial information about the state of the system, being informed only about the total number of customers in the network. After receiving this information, they decide whether to join or balk the system.

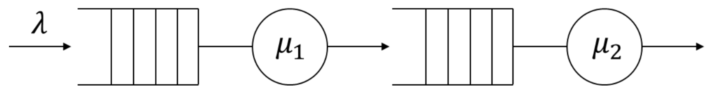


Figure 2.1: Two-node tandem network with arrival process given by a Poisson process with rate λ and service times with rates μ_1 and μ_2 at each corresponding node.

We suppose that a tagged customer gets a reward R for joining the network and receiving the service, and pays for each unit of sojourn time at node l a cost C_l with a resulting random profit given by

$$P = R - C_1 S_1 - C_2 S_2,$$

where S_l denotes the sojourn time she would spend at node l.

The tagged user makes her decision by optimizing the expected profit given the information she receives at her arrival. That is, assuming that she is informed that the total number of customers in the network before her arrival is k, she computes her expected profit as

$$P(k) = R - C_1 T_1(k) - C_2 T_2(k),$$

where $T_l(k)$ is her expected sojourn time at node l given the information she received. We assume that

$$R \geq \frac{C_1}{\mu_1} + \frac{C_2}{\mu_2},$$

since, otherwise, even a customer finding an empty system would trivially decide to balk.

The aim of this chapter is to find an equilibrium threshold strategy for this model, under the partial information, such that the expected profit of any customer is maximized. We assume that all customers are using the same pure threshold strategy and we denote by K the equilibrium threshold. Hence, all users join the network if and only if it contains less than K customers. We will see that in a two-node tandem network the pure threshold K exists and is defined such that

$$P(k) \geq 0 \quad \text{as} \quad k < K,$$
$$P(k) < 0 \quad \text{as} \quad k \geq K.$$

2.2 Expected sojourn times under total information

We consider the state space $\mathbb{N}^2$, which is the set of all possible pairs (q_1, q_2) with q_l being the length of the queue at node l, $l = 1, 2$. In this section, we analyze the expected sojourn time under total information, that is knowing the complete state of the system, say $(q_1, q_2) = (n-1, m)$, before deciding whether to enter the system or not.

Let $S_l(n,m)$ be the sojourn time spent at queue l by a customer that joins a system being in state $(n-1,m)$, that is she is going to occupy position n in the first queue, and $T_l(n,m) = \mathbb{E}[S_l(n,m)]$ be the corresponding expected sojourn time with $T(n,m) = T_1(n,m) + T_2(n,m)$ the total expected sojourn time.

The following results characterize the expected sojourn times.

Lemma 2.1 (Mean sojourn time at the first queue)**.** The expected sojourn time spent at the first queue, $T_1(n,m)$, is

$$T_1(n,m) = \frac{n}{\mu_1}, \quad n \geq 0. \tag{2.1}$$

Proof: Since the first queue has a service rate μ_1, then $S_1(n,m)$ is sum of n independent exponential random variable with mean μ_1^{-1}, that is $S_1(n,m) \sim \text{Erlang}(n, \mu_1)$. So we have that

$$T_1(n,m) = \mathbb{E}[S_1(n,m)] = \frac{n}{\mu_1}.$$

□

Lemma 2.2 (Mean sojourn time)**.** The total expected sojourn time, $T(n,m)$, is defined by the following recursive formula

$$T(n,m) = \frac{1}{\mu_1+\mu_2} + \frac{\mu_1}{\mu_1+\mu_2} T(n-1, m+1) + \frac{\mu_2}{\mu_1+\mu_2} T(n, m-1), \quad n, m > 0, \tag{2.2}$$

and the boundary conditions

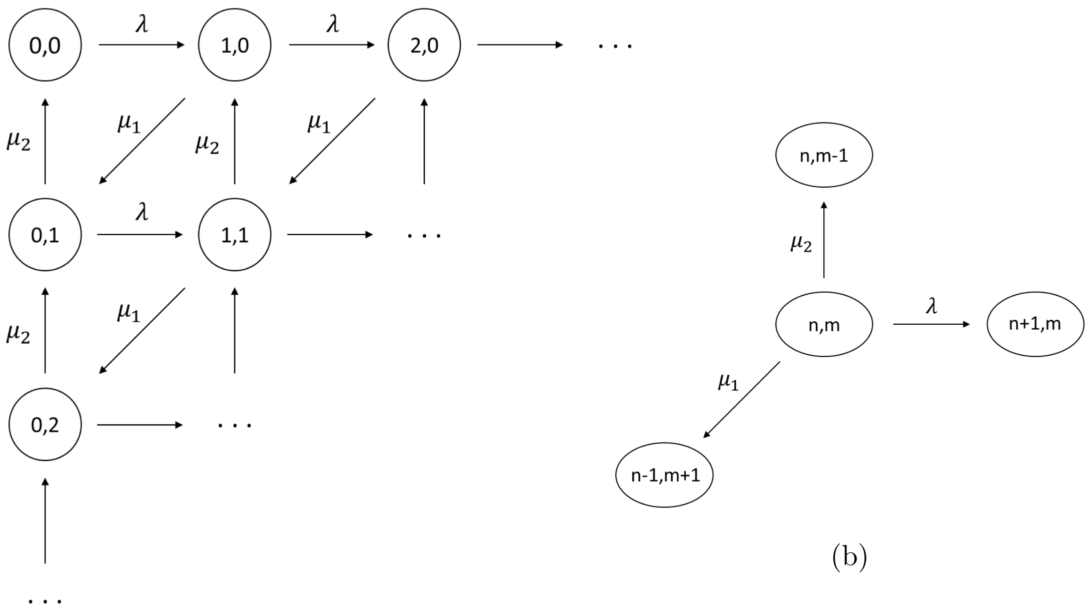
$$\begin{aligned} T(0,m) &= \frac{m}{\mu_2}, \quad m \geq 0, && (2.3) \\ T(n+1,0) &= \frac{1}{\mu_1} + T(n,1), \quad n \geq 0. && (2.4) \end{aligned}$$

Proof: To compute the total expected sojourn time, we consider the possible states of the system, that are described in the graphs of Figure 2.2.

Now, we suppose to follow the tagged customer in position (n,m) until she reaches the position $(0,0)$ and compute how long she takes on average, that is $T(n,m)$. So we don't consider the possibility of going from state (n,m) to state $(n+1,m)$, because the customer cannnot be overtaken by those after her. It means that only those in front of her can affect her sojourn time.

So the position of the tagged customer in the system can be represented by a uniform Markov chain

$$(Q(t))_{t\geq 0} = \{(Q_1(t), Q_2(t)) : t \geq 0\},$$



(a) (b)

Figure 2.2: Transition rate diagrams for the uniform Markov chain describing the dynamics of the system, where μ_l denotes the service rate at node $l = 1, 2$ and λ denotes the arrival rate.

with state space $\mathbb{N}^2$, where $Q_l(t)$ represents the number of customers at node l at time t. The uniform Markov chain $(Q(t))_{t\geq 0}$ is defined as follows

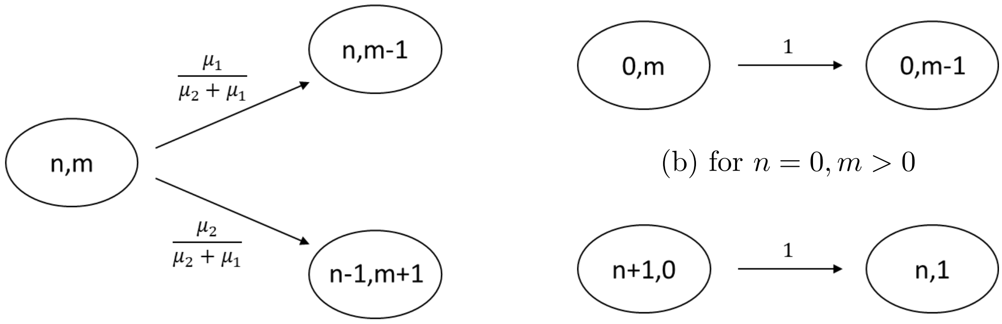
$$Q(t) = \hat{Q}_{N(t)},$$

where $(\hat{Q}_n)_{n\geq 0}$ is a discrete-time homogeneous Markov chain with values in $\mathbb{N}^2$, $\hat{Q}_0 = (n, m)$ and the transition probabilities are those described in Figure 2.3; $N(t)$ is the counting process associated to a Poisson process with intensity

$$\begin{cases} \mu_1 + \mu_2 & \text{if} \quad n, m > 0 \\ \mu_2 & \text{if} \quad n = 0, m > 0 \\ \mu_1 & \text{if} \quad n \geq 0, m = 0 \end{cases}.$$

Therefore, applying a first step analysis to this uniform Markov chain for $n, m > 0$, the total expected sojourn time can be computed recursively with the following formula

$$T(n, m) = \frac{1}{\mu_1 + \mu_2} + \frac{\mu_1}{\mu_1 + \mu_2} T(n - 1, m + 1) + \frac{\mu_2}{\mu_1 + \mu_2} T(n, m - 1), \quad n, m > 0,$$



(a) for $n, m > 0$

(b) for $n = 0, m > 0$

(c) for $n \geq 0, m = 0$

Figure 2.3: Transition diagram for the embedded discrete-time Markov chain. Case (a) is for $n, m > 0$, cases (b) and (c) refer to the boundary conditions.

and, similarly, we obtain the boundary conditions

$$T(0, m) = \frac{m}{\mu_2}, \quad m \geq 0;$$
$$T(n+1, 0) = \frac{1}{\mu_1} + T(n, 1), \quad n \geq 0.$$

□

Lemma 2.3 (Mean sojourn time at the second queue)**.** The expected sojourn time spent at the second queue, $T_2(n, m)$, can be computed with the following recursive formula

$$T_2(n, m) = \left(\frac{\mu_2}{\mu_1 + \mu_2}\right)^m T_2(n-1, 1) + \frac{\mu_1}{\mu_1 + \mu_2} \sum_{k=0}^{m-1} \left(\frac{\mu_2}{\mu_1 + \mu_2}\right)^k T_2(n-1, m+1-k), \quad n > 0, m \geq 0, \tag{2.5}$$

and the boundary condition

$$T_2(0, m) = \frac{m}{\mu_2}, \quad m \geq 0.$$

Proof: Knowing that $T(n, m) = T_1(n, m) + T_2(n, m)$, using Eq. (2.2) and Eq. (2.1), we get the following recursive formula for $T_2(n, m)$

$$T_2(n, m) = \frac{\mu_1}{\mu_1 + \mu_2} T_2(n-1, m+1) + \frac{\mu_2}{\mu_1 + \mu_2} T_2(n, m-1), \quad n, m > 0. \tag{2.6}$$

The boundary condition is immediately recovered from Eq. (2.3) because

$$\begin{aligned} & T(0,m) = \frac{m}{\mu_2} \\ \Leftrightarrow \quad & T_1(0,m) + T_2(0,m) = \frac{m}{\mu_2} \\ \Leftrightarrow \quad & T_2(0,m) = \frac{m}{\mu_2}. \end{aligned}$$

By an induction argument, we can prove Eq. (2.5): for $m = 0$, we want to prove $T_2(n,0) = T_2(n-1,0)$ and, from (2.4), we have that

$$\begin{aligned} & T(n,0) = \frac{m}{\mu_1} + T(n-1,1) \\ \Leftrightarrow \quad & T_1(n,0) + T_2(n,0) = \frac{m}{\mu_1} + T_1(n-1,1) + T_2(n-1,1) \\ \Leftrightarrow \quad & T_2(n,0) = T_2(n-1,1). \end{aligned}$$

Now, by inductive hypothesis, let (2.5) be true for $m \geq 0$ and we prove it for $m+1$: from (2.6) we have that

$$\begin{aligned} T_2(n,m+1) &= \frac{\mu_1}{\mu} T_2(n-1,m+2) + \frac{\mu_2}{\mu} T_2(n,m) \\ &= \frac{\mu_1}{\mu} T_2(n-1,m+2) + \frac{\mu_2}{\mu}\left[\left(\frac{\mu_2}{\mu}\right)^m T_2(n-1,1) + \frac{\mu_1}{\mu}\sum_{k=0}^{m-1}\left(\frac{\mu_2}{\mu}\right)^k T_2(n-1,m+1-k)\right] \\ &= \left(\frac{\mu_2}{\mu}\right)^{m+1} T_2(n-1,1) + \left[T_2(n-1,m+2) + \sum_{k=0}^{m-1}\left(\frac{\mu_2}{\mu}\right)^{k+1} T_2(n-1,m+1-k)\right] \\ &= \left(\frac{\mu_2}{\mu}\right)^{m+1} T_2(n-1,1) + \frac{\mu_1}{\mu_1+\mu_2}\sum_{k=0}^{m}\left(\frac{\mu_2}{\mu}\right)^k T_2(n-1,m+2-k), \end{aligned}$$

where $\mu = \mu_1 + \mu_2$. □

The following lemma shows the conditions under which the functions $T(n,m)$ and $T_l(n,m)$ are monotone in the variable n.

Lemma 2.4 (Monotonicity of the mean sojourn times)**.** The functions $T_1(n,m)$ and $T(n,k-n)$ are increasing in n. The function $T_2(n,m)$ is non decreasing in n if and only if $\mu_1 \geq \mu_2$.

Proof: By Lemma 2.1, we have that

$$T_1(n,m) = \frac{n}{\mu_1},$$

so $T_1(n,m)$ is increasing in n, and the same holds for $T_1(n,k-n)$, where $k = n+m$.

To prove that $T(n, k-n)$ is increasing in n, with $k = n+m$, we prove by induction that

$$T(n+1, m) > T(n, m+1), \quad \forall n, m > 0.$$

For $m = 0$, we want to prove that $T(n+1, 0) > T(n, 1)$ and we have that from (2.4)

$$T(n+1, 0) = \frac{1}{\mu_1} + T(n, 1) > T(n, 1).$$

For $n = 0$, we want to prove that $T(1, m) > T(0, m+1)$ and we do it by an inductive argument: for $m = 0$ from (2.4)

$$T(1, 0) = \frac{1}{\mu_1} + T(0, 1) > T(0, 1).$$

Let $T(1, m-1) > T(0, m)$ be true, then from (2.2)

$$\begin{aligned} T(1, m) &= \frac{1}{\mu} + \frac{\mu_1}{\mu} T(0, m+1) + \frac{\mu_2}{\mu} T(1, m-1) \\ &> \frac{1}{\mu} + \frac{\mu_1}{\mu} T(0, m+1) + \frac{\mu_2}{\mu} T(0, m) \\ &= \frac{\mu_2}{\mu} \left(\frac{1}{\mu_2} + T(1, m) \right) + \frac{\mu_1}{\mu} T(0, m+1) \\ &= \frac{\mu_2}{\mu} T(0, m+1) + \frac{\mu_1}{\mu} T(0, m+1) = T(0, m+1). \end{aligned}$$

To use the induction procedure, we firstly define a partial order: we say that $(n_1, m_1) \prec (n_2, m_2)$, if $(n_1 < n_2)$ or $(n_1 = n_2 \ \& \ m_1 < m_2)$. Now, we prove that $T(n+1, m) > T(n, m+1)$, knowing that the same hold for any $(\bar{n}, \bar{m}) \prec (n, m)$: from (2.2)

$$\begin{aligned} T(n+1, m) &= \frac{1}{\mu} + \frac{\mu_1}{\mu} T(n, m+1) + \frac{\mu_2}{\mu} T(n+1, m-1) \\ &> \frac{1}{\mu} + \frac{\mu_1}{\mu} T(n-1, m+2) + \frac{\mu_2}{\mu} T(n, m) \\ &= T(n, m+1), \end{aligned}$$

where we use the inequalities $T(n+1, m-1) > T(n, m)$ and $T(n, m+1) > T(n-1, m+2)$ (because $(n, m-1) \prec (n, m)$ and $(n-1, m+1) \prec (n, m)$).

Finally, to show that $T_2(n, m)$ is non decreasing in n if $\mu_1 \geq \mu_2$, we consider $\Delta_1 T_2(n, m) = T_2(n+1, m) - T_2(n, m)$ and we prove that $\Delta_1 T_2(n, m) \geq 0$. From (2.5) we obtain that for $n > 0, m \geq 0$

$$\begin{aligned} \Delta_1 T_2(n, m) = & \left(\frac{\mu_2}{\mu_1 + \mu_2} \right)^m \Delta_1 T_2(n-1, 1) \\ & + \frac{\mu_1}{\mu_1 + \mu_2} \sum_{k=0}^{m-1} \left(\frac{\mu_2}{\mu_1 + \mu_2} \right)^k \Delta_1 T_2(n-1, m+1-k). \end{aligned}$$

Using an induction argument on n, it is sufficient to prove the base case $\Delta_1 T_2(0,m) \geq 0$, because the induction step then follows as all the coefficients in (2.2) are positive. After some computations, we obtain that

$$\begin{aligned}
\Delta_1 T_2(0,m) &= T_2(1,m) - T_2(0,m) \\
&= \left(\frac{\mu_2}{\mu}\right)^m T_2(0,1) + \frac{\mu_1}{\mu}\sum_{k=0}^{m-1}\left(\frac{\mu_2}{\mu}\right)^k T_2(0,m+1-k) - T_2(0,m) \\
&= \frac{1}{\mu_2}\left(\frac{\alpha - 1 + (\alpha+1)^{-m}}{\alpha}\right),
\end{aligned}$$

where $\alpha = \mu_1/\mu_2$. It holds that $\Delta_1 T_2(0,m)$ is decreasing in m and

$$\lim_{m\to\infty} \Delta_1 T_2(0,m) = \frac{1}{\mu_2}\left(\frac{\alpha-1}{\alpha}\right) \geq 0 \quad \Leftrightarrow \quad \alpha \geq 1,$$

So, $\mu_1 \geq \mu_2$ if and only if $\Delta_1 T_2(0,m) \geq 0$ for all $m \geq 0$, and this concludes the proof. □

2.3 Expected sojourn times under partial information

In this section, we analyze the expected sojourn time under partial information, that is only knowing the total number of customers in the system, say k.

So, we assume that all arriving customers receive a partial information about the state of the network and decide to join if and only if the number of customers k in the system is less than a given threshold $K \geq 0$. Under the K-strategy the tandem network behaves as a semiopen Jackson network (Section 1.1), with an overall buffer capacity of K, where we have two nodes and routing probabilities $p_{01} = 1$, $p_{12} = 1$, $p_{20} = 1$. From the traffic equation

$$\lambda_i = \lambda p_{0i} + \sum_{j=1}^{2} \lambda_j p_{ji}, \quad i = 1,2,$$

we obtain that $\lambda_i = \lambda$ for $i = 1,2$.

Let Q_l^* be the stationary number of customers at node l and $Q^* = Q_1^* + Q_2^*$ be the stationary total number of customers in the system.

Let $\pi_K(n,m) = \mathbb{P}_K(Q_1^* = n, Q_2^* = m)$ be the stationary distribution of the state of the system, conditioned on $Q^* \leq K$, and $p_l(n \mid k) = \mathbb{P}_K(Q_l^* = n \mid Q^* = k)$ the distribution of queue lengths, knowing the partial information $Q^* = k$. The following lemma gives an expression for these two distributions.

Lemma 2.5 (Stationary distribution)**.** The stationary distribution is given by

$$\pi_K(n,m) = c_K \rho_1^n \rho_2^m, \quad n+m \le K,$$

where $\rho_l = \lambda/\mu_l$ and $c_K^{-1} = \sum_{n+m\le K} \rho_1^n \rho_2^m$ is the normalization constant. Furthermore, it holds that

$$p_l(n \mid k) = \begin{cases} \mu_l^{k-n}\mu_{3-l}^n(\mu_1-\mu_2)/(\mu_1^{k+1}-\mu_2^{k+1}), & \mu_1 \ne \mu_2 \\ 1/(1+k), & \mu_1 = \mu_2 \end{cases}. \tag{2.7}$$

Proof: Since the tandem network is supposed to be a semiopen Jackson model, using Theorem 1.3, the stationary distribution is:

$$\begin{aligned} \pi_K(n,m) &= \frac{\mathbb{P}(Y_1=n)\mathbb{P}(Y_2=m)}{\mathbb{P}(Y_1+Y_2\le K)} \\ &= \frac{\mathbb{P}(Y_1=0)(\lambda_1^n/\mu_1^n)\,\mathbb{P}(Y_2=0)(\lambda_2^n/\mu_2^n)}{\sum_{n+m\le K}\mathbb{P}(Y_1=n)\mathbb{P}(Y_2=m)} \\ &= \frac{\mathbb{P}(Y_1=0)\rho_1^n\,\mathbb{P}(Y_2=0)\rho_2^n}{\mathbb{P}(Y_1=0)\mathbb{P}(Y_2=0)\sum_{n+m\le K}\rho_1^n\rho_2^m} = \frac{\rho_1^n\rho_2^m}{\sum_{n+m\le K}\rho_1^n\rho_2^m}, \end{aligned}$$

where $Y=(Y_1,Y_2)$ is defined in Eq. (1.4).

Assuming $n \le K$, the conditional probability is

$$\begin{aligned} \mathbb{P}_K(Q_l^*=n \mid Q^*=k) &= \frac{\mathbb{P}_K(Q_l^*=n, Q^*=k)}{\mathbb{P}_K(Q^*=k)} \\ &= \frac{\mathbb{P}_K(Q_l^*=n, Q_{3-l}^*=k-n)}{\sum_{h=0}^k \mathbb{P}_K(Q_l^*=h, Q_{3-l}^*=k-h)} \\ &= \frac{c_K\rho_l^n\rho_{3-l}^{k-n}}{\sum_{h=0}^k c_K\rho_l^h\rho_{3-l}^{k-h}} = \frac{\rho_l^n\rho_{3-l}^{k-n}}{\sum_{h=0}^k \rho_l^h\rho_{3-l}^{k-h}}. \end{aligned}$$

After some algebraic manipulation, we obtain Eq. (2.7): if $\mu_1=\mu_2$, then $\rho_1=\rho_2$ and

$$p_l(n \mid k) = \frac{\rho_l^n\rho_{3-l}^{k-n}}{\sum_{h=0}^k \rho_l^h\rho_{3-l}^{k-h}} = \frac{\rho_l^k}{\sum_{h=0}^k \rho_l^k} = \frac{1}{1+k},$$

while, if $\mu_1 \ne \mu_2$

$$\begin{aligned} p_l(n \mid k) &= \frac{\rho_l^n\rho_{3-l}^{k-n}}{\sum_{h=0}^k \rho_l^h\rho_{3-l}^{k-h}} = \frac{(\mu_l^n\mu_{3-l}^{k-n})^{-1}}{\sum_{h=0}^k(\mu_l^h\mu_{3-l}^{k-h})^{-1}} = \frac{(\mu_{3-l}/\mu_l)^n}{\sum_{h=0}^k(\mu_{3-l}/\mu_l)^h} \\ &= \frac{(\mu_{3-l}/\mu_l)^n(1-\mu_{3-l}/\mu_l)}{1-(\mu_{3-l}/\mu_l)^{k+1}} = \frac{\mu_l^{k-n}\mu_{3-l}^n(\mu_l-\mu_{3-l})}{\mu_l^{k+1}-\mu_{3-l}^{k+1}} \\ &= \frac{\mu_l^{k-n}\mu_{3-l}^n(\mu_1-\mu_2)}{\mu_1^{k+1}-\mu_2^{k+1}}, \end{aligned}$$

where in the last step we substitute l and $3-l$, with 1 and 2, respectively, where possible. □

We observe that the conditional probability $p_l(n|k)$ doesn't depend on K and nor on the arrival rate λ. So, it means that it is sufficient to know that the arrival process is a Poisson process, but we are not interested in knowing its intensity.

Let us define by $T_l(k) = \mathbb{E}[S_l|Q^* = k]$ the expected sojourn time at queue l of a tagged customer that enters the system containing k customers. In the next lemma, we show an explicit formula for $T_l(k)$ and $T(k)$.

Lemma 2.6 (Expected sojourn times under partial information)**.** For $\mu_1 \neq \mu_2$, it holds that

$$T_l(k) = \frac{1}{\mu_l - \mu_{3-l}} - \frac{k+1}{\mu_l}\frac{\mu_{3-l}^{k+1}}{\mu_l^{k+1} - \mu_{3-l}^{k+1}}, \quad l = 1,2, \tag{2.8}$$

$$T(k) = \frac{k+1}{\mu_1\mu_2}\frac{\mu_1^{k+2} - \mu_2^{k+2}}{\mu_1^{k+1} - \mu_2^{k+1}}. \tag{2.9}$$

For $\mu_1 = \mu_2$, it holds that

$$\begin{aligned} T_l(k) &= \frac{1}{\mu_1}\left(1 + \frac{k}{2}\right), \quad l = 1,2, \\ T(k) &= \frac{2}{\mu_1}\left(1 + \frac{k}{2}\right). \end{aligned} \tag{2.10}$$

Proof: By definition and using Eq. (2.1), we have that

$$T_1(k) = \sum_{n=0}^{k} T_1(n+1, k-n)p_1(n|k) = \frac{1}{\mu_1}\sum_{n=0}^{k}(n+1)p_1(n|k).$$

So, using (2.7), we get that, for $\mu_1 \neq \mu_2$

$$\begin{aligned} T_1(k) &= \frac{1}{\mu_1}\frac{\mu_1 - \mu_2}{\mu_1^{k+1} - \mu_2^{k+1}}\sum_{n=0}^{k}(n+1)\mu_1^{k-n}\mu_2^{n} \\ &= \frac{1}{\mu_1}\frac{\mu_1 - \mu_2}{\mu_1^{k+1} - \mu_2^{k+1}}\frac{\mu_1^{2+k} - (2+k)\mu_1\mu_2^{k+1} + (k+1)\mu_2^{2+k}}{(\mu_1 - \mu_2)^2} \\ &= \frac{\mu_1(\mu_1^{k+1} - \mu_2^{k+1}) - (k+1)\mu_2^{k+1}(\mu_1 - \mu_2)}{\mu_1(\mu_1^{k+1} - \mu_2^{k+1})(\mu_1 - \mu_2)} \\ &= \frac{1}{\mu_1 - \mu_2} - \frac{k+1}{\mu_1}\frac{\mu_2^{k+1}}{\mu_1^{k+1} - \mu_2^{k+1}}. \end{aligned}$$

Regarding $T(k)$, by definition it holds that

$$T(k) = \sum_{n=0}^{k} T(n+1, k-n)p_1(n|k),$$

and using (2.7), we get that, for $\mu_1 \neq \mu_2$

$$T(k) = \frac{\mu_1 - \mu_2}{\mu_1^{k+1} - \mu_2^{k+1}} \sum_{n=0}^{k} T(n+1, k-n) \mu_1^{k-n} \mu_2^n$$
$$= \left(1 - \frac{\mu_2}{\mu_1}\right) \frac{\mu_1^{k+1}}{\mu_1^{k+1} - \mu_2^{k+1}} \sum_{n=0}^{k} T(n+1, k-n) \left(\frac{\mu_2}{\mu_1}\right)^n.$$

So we have that proving (2.9) is equivalent to proving

$$\sum_{n=0}^{k} T(n+1, k-n) \left(\frac{\mu_2}{\mu_1}\right)^n = \frac{(k+1)(\mu_1^{k+2} - \mu_2^{k+2})}{\mu_2(\mu_1 - \mu_2)\mu_1^{k+1}}. \tag{2.11}$$

We prove (2.11) by induction on k. The base case is for $k = 0$:

$$T(1,0) = \frac{\mu_1 + \mu_2}{\mu_2 \mu_1} = \frac{1}{\mu_1} + \frac{1}{\mu_2},$$

and this is true because $T(1,0) = 1/\mu_1 + T(0,1) = 1/\mu_1 + 1/\mu_2$.

Let Eq. (2.11) be true for $k-1$, that is

$$\sum_{n=0}^{k-1} T(n+1, k-1-n) \left(\frac{\mu_2}{\mu_1}\right)^n = \frac{k(\mu_1^{k+1} - \mu_2^{k+1})}{\mu_2(\mu_1 - \mu_2)\mu_1^k},$$

then

$$\sum_{n=0}^{k} T(n+1, k-n) \left(\frac{\mu_2}{\mu_1}\right)^n = \sum_{n=0}^{k-1} T(n+1, k-n) \left(\frac{\mu_2}{\mu_1}\right)^n + T(k+1, 0) \left(\frac{\mu_2}{\mu_1}\right)^k$$
$$= \sum_{n=0}^{k-1} \frac{1}{\mu_1 + \mu_2} \left(\frac{\mu_2}{\mu_1}\right)^n + \sum_{n=0}^{k-1} \frac{\mu_1}{\mu_1 + \mu_2} T(n, k-n+1) \left(\frac{\mu_2}{\mu_1}\right)^n$$
$$+ \sum_{n=0}^{k-1} \frac{\mu_2}{\mu_1 + \mu_2} T(n+1, k-n-1) \left(\frac{\mu_2}{\mu_1}\right)^n + T(k+1, 0) \left(\frac{\mu_2}{\mu_1}\right)^k$$
$$= \sum_{n=0}^{k-1} \frac{1}{\mu_1 + \mu_2} \left(\frac{\mu_2}{\mu_1}\right)^n + \frac{\mu_2}{\mu_1 + \mu_2} \frac{k(\mu_1^{k+1} - \mu_2^{k+1})}{\mu_2(\mu_1 - \mu_2)\mu_1^k} + \frac{\mu_1}{\mu_1 + \mu_2} T(0, k+1)$$
$$+ \frac{1}{\mu_1 + \mu_2} \left(\frac{\mu_2}{\mu_1}\right)^k + \frac{\mu_1}{\mu_1 + \mu_2} \sum_{n=0}^{k} T(n+1, k-n) \left(\frac{\mu_2}{\mu_1}\right)^n$$
$$= \frac{\mu_1}{\mu_1 + \mu_2} \left(\frac{(k+1)(\mu_1^{k+2} - \mu_2^{k+2})}{\mu_2(\mu_1 - \mu_2)\mu_1^{k+1}}\right) + \frac{\mu_1}{\mu_1 + \mu_2} \sum_{n=0}^{k} T(n+1, k-n) \left(\frac{\mu_2}{\mu_1}\right)^n,$$

where we used in the second step the recursive formula (2.2) for $T(n+1, k-n)$, in the third step the inductive hypothesis and the boundary conditions (2.4) and in the fourth step the boundary conditions (2.3). Finally, rearranging terms, we get (2.11).

To compute $T_2(k)$, we observe that

$$\begin{aligned} T(k) &= \sum_{n=0}^{k} T(n+1, k-n) p_1(n|k) \\ &= \sum_{n=0}^{k} T_1(n+1, k-n) p_1(n|k) + \sum_{n=0}^{k} T_2(n+1, k-n) p_1(n|k) \\ &= T_1(k) + T_2(k), \end{aligned}$$

therefore, we get that, for $\mu_1 \neq \mu_2$,

$$T_2(k) = T(k) - T_1(k) = \frac{1}{\mu_2 - \mu_1} - \frac{k+1}{\mu_2} \frac{\mu_1^{k+1}}{\mu_2^{k+1} - \mu_1^{k+1}}.$$

For $\mu_1 = \mu_2$, we compute the expression of $T_l(k)$ as the limit for the expression (2.8) as $\mu_1 \to \mu_2$,

$$\begin{aligned} T_l(k) &= \lim_{\mu_1 \to \mu_2} \left(\frac{1}{\mu_l - \mu_{3-l}} - \frac{k+1}{\mu_l} \frac{\mu_{3-l}^{k+1}}{\mu_l^{k+1} - \mu_{3-l}^{k+1}} \right) \\ &= \lim_{x \to 1} \frac{1}{\mu_1} \left(\frac{1}{x-1} - \frac{k+1}{x(x^{k+1}-1)} \right) = \frac{1}{\mu_1} \left(1 + \frac{k}{2} \right) \end{aligned}$$

and $T(k)$ results to be $T_1(k) + T_2(k)$. □

2.4 Equilibrium strategies

By Lemma 2.6, the expected profit of the tagged customer

$$P(k) = R - C_1 T_1(k) - C_2 T_2(k)$$

does not depend on the threshold K, which characterizes the strategy employed by all customers.

The tagged customer decides to enter only if $P(k) \geq 0$. In the sequel we show that the expected profit function is decreasing in k and we obtain that the equilibrium strategy is a dominant strategy in the class of threshold strategies, using the fact that $P(k)$ does not depend on K.

First, we prove that $T_l(k)$ is increasing in k. To do this, we could use some analysis results or computational methods, but we prefer using some probability arguments. The following lemma is about the stochastic monotonicity of the random variables $Q_l^*(k)$, which represent the stationary random number of customers at queue l under the partial information.

Lemma 2.7 (Stochastic monotonicity of $Q_l^*(k)$)**.** The random variables $Q_l^*(k)$ are increasing stochastically ordered in $k \geq 0$, that is

$$Q_l^*(k+1) \geq_{st} Q_l^*(k).$$

Proof: In order to show that $Q_l^*(k+1) \geq_{st} Q_l^*(k)$, i.e.

$$\mathbb{P}(Q_l^*(k+1) \geq n) \geq \mathbb{P}(Q_l^*(k) \geq n),$$

it is enough to prove, by Theorem 1.7, the stronger condition given by the likelihood ratio ordering $Q_l^*(k+1) \geq_{\ell r} Q_l^*(k)$, i.e.

$$\mathbb{P}(Q_l^*(k+1) = n+1)\mathbb{P}(Q_l^*(k) = n) \geq \mathbb{P}(Q_l^*(k+1) = n)\mathbb{P}(Q_l^*(k) = n+1).$$

To prove this, we have that for $n < k$

$$\begin{aligned} \mathbb{P}(Q_l^*(k+1) = n+1)\mathbb{P}(Q_l^*(k) = n) &= \frac{p_l^{n+1} p_{3-l}^{k+1-(n+1)}}{\sum_{h=0}^{k+1} p_l^h p_{3-l}^{k+1-h}} \frac{p_l^n p_{3-l}^{k-n}}{\sum_{h=0}^{k} p_l^h p_{3-l}^{k-h}} \\ = \frac{p_l^n p_{3-l}^{k+1-n}}{\sum_{h=0}^{k+1} p_l^h p_{3-l}^{k+1-h}} \frac{p_l^{n+1} p_{3-l}^{k-(n+1)}}{\sum_{h=0}^{k} p_l^h p_{3-l}^{k-h}} &= \mathbb{P}(Q_l^*(k+1) = n)\mathbb{P}(Q_l^*(k) = n+1), \end{aligned}$$

and for $n = k$

$$\mathbb{P}(Q_l^*(k+1) = k+1)\mathbb{P}(Q_l^*(k) = k) \geq 0 = \mathbb{P}(Q_l^*(k+1) = k)\mathbb{P}(Q_l^*(k) = k+1).$$

□

Lemma 2.7 implies the following monotonicity result for the mean sojourn functions.

Lemma 2.8 (Monotonicity of the expected sojourn times)**.** The functions $T_l(k)$ and $T(k)$ are increasing in k.

Proof: We recall the following result (*Remark* 1.5): for two random variables X and Y

$$X \geq_{st} Y \quad \Leftrightarrow \quad \forall u \quad \text{non decreasing function} \quad \mathbb{E}[u(X)] \geq \mathbb{E}[u(Y)].$$

Using this result, Lemma 2.4 and Lemma 2.7, we get that

$$\begin{aligned} T_1(k+1) &= \mathbb{E}[T_1(Q_1^*(k+1), k+1-Q_1^*(k+1))] \\ &= \mathbb{E}[T_1(Q_1^*(k+1), k-Q_1^*(k+1))] \\ &\geq \mathbb{E}[T_1(Q_1^*(k), k-Q_1^*(k))] = T_1(k), \end{aligned}$$

where in the inequality we use the fact that $Q_1^*(k+1) \geq_{st} Q_1^*(k)$ and $T_1(n, k-n)$ is non decreasing in n. Then, since $T(k+1) \neq T(k)$, we have that $T(k+1) > T(k)$. By the symmetry of the formulas of $T_1(k)$ and $T_2(k)$ (Eq. (2.8)), we obtain directly that also $T_2(k)$ is increasing in k. The same holds for $T(k)$, because $T(k) = T_1(k) + T_2(k)$. □

From the previous Lemma 2.8, we have that the expected profit function $P(k)$ is decreasing for all values of μ_1 and μ_2, since $P(k)$ is defined as in Eq. (2.4).

Finally, we state the main result that gives the threshold strategy to be adopted in order to have an equilibrium.

Theorem 2.9 (Equilibrium threshold strategy)**.** When the information known by the arriving customers is the total number of customers in the network only, then the equilibrium threshold strategy is given by the threshold K such that

$$K = \inf\{k \in \mathbb{Z}_+ : P(k) < 0\}. \tag{2.12}$$

According to this strategy, the tagged customer enters only if she finds less than K customers in the system. The K-strategy is a *dominant strategy* in the class of threshold strategies.

Proof: The system is ergodic, which means there is a positive probability such that the system eventually is in any state. So, without losing of generality, we can assume it starts empty. In particular, we note that the system is empty with probability

$$\mathbb{P}_K(Q^* = 0) = \mathbb{P}_K(Q_1^* = 0, Q_2^* = 0) = c_K > 0,$$

and, for $k = 0$, the expected profit function is

$$P(0) = R - \frac{C_1}{\mu_1} - \frac{C_2}{\mu_2},$$

which is supposed to be positive.

Let K be the index obtained by Eq. (2.12), including $K = \infty$. We show that the strategy K is a dominant strategy by proving that it is the best response to any other possible threshold strategy.

Indeed, we consider $F_{K_1,K_2}(k)$ as the payoff of a customer that chooses to follow the K_1-strategy, when everyone else chooses to follow the K_2-strategy. In this case, if $K_1 \leq K_2$, the payoff function is

$$F_{K_1,K_2}(k) = \begin{cases} P(k), & \text{if} \quad 0 \leq k < K_1 \\ 0, & \text{if} \quad K_1 \leq k \leq K_2 \end{cases},$$

because the customer receives a payoff $P(k)$ if she enters the system, 0 otherwise. Let us note that, since all customers employ the K_2-strategy, the tagged user will never find more than K_2 customers in the system. If, instead, $K_1 > K_2$, then the payoff function is

$$F_{K_1,K_2}(k) = P(k), \quad \text{for any} \quad 0 \leq k \leq K_2$$

From Definition 1.18, the K-strategy is a dominant strategy, if

$$F_{K,K_2}(k) \geq F_{K_1,K_2}(k) \quad \forall K_1 \in \mathbb{N}, \quad \forall K_2 \in \mathbb{N},$$

and this holds because of the definition of F. □

We observed that Theorem 2.9 is saying that, if all other customers are using the same K_2-strategy for any K_2, then for a tagged customer, that is deciding whether to join or not the system, the best strategy to follow is the K-strategy.

The K-strategy of Theorem 2.9 is a pure threshold strategy, and, it is the only equilibrium strategy for this model, if $P(K-1) > 0$. If, instead, $P(K-1) = 0$, then the equilibrium strategies are the K-strategy and the $(K-1)$-strategy, which are pure threshold strategies, and all the strictly convex combinations between these two strategies, which turn out to be mixed threshold strategy. There are no other threshold strategies for this model, but there are other non-threshold strategies. The same reasoning of *Remark* 1.22 holds also for this model.

We observe that for $\mu_1 = \mu_2$, by Eq. (2.10), the expected profit function is

$$P(k) = R - \frac{C_1 + C_2}{\mu_1}\left(1 + \frac{k}{2}\right)$$

and the equilibrium threshold strategy is

$$K = \left\lfloor \frac{2R\mu_1}{C_1 + C_2} \right\rfloor - 1. \tag{2.13}$$

Assuming also that $C_1 = C_2$, we have that

$$K = \left\lfloor \frac{R\mu_1}{C_1} \right\rfloor - 1, \tag{2.14}$$

which is equal to (1.6) a part from a -1.

Now, we analyze the situation for μ_l goes to infinity, for $l = 1, 2$. We expect that the model becomes equal to a model with one single queue with service rate μ_{3-l} and cost to pay for staying in the system C_{3-l}. Indeed, as soon as a customer reaches queue l, she immediately moves to queue $3-l$, if $l = 1$, or she leaves the system, if $l = 2$. The result is that queue l does not affect the sojourn of a customer in the system. We can prove this result also by making the limit of $T_l(k), T_{3-l}(k)$ and $P(k)$. We obtain that

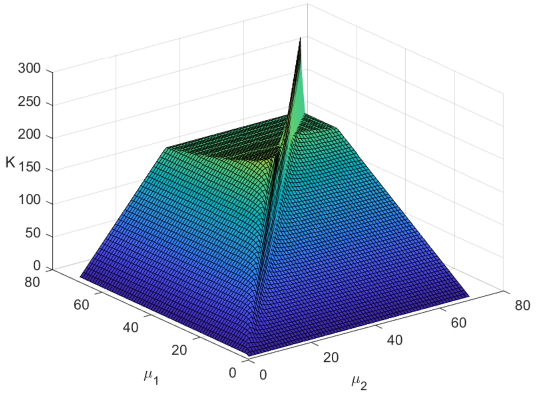
$$\lim_{\mu_l \to \infty} T_l(k) = 0,$$
$$\lim_{\mu_l \to \infty} T_{3-l}(k) = \frac{k+1}{\mu_{3-l}},$$
$$\lim_{\mu_l \to \infty} P(k) = R - C_{3-l}\frac{k+1}{\mu_{3-l}},$$

and these quantities are the same as for the observable $M/M/1$ model (Section 1.5).

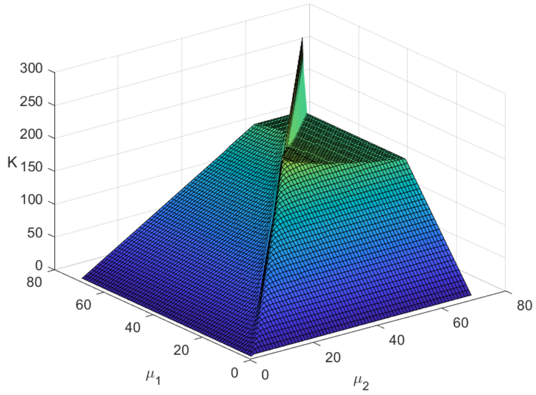
2.5 Numerical computations

For some numerical computations of the equilibrium threshold K, we look at the two-node tandem network for different values of C_1 and C_2, as the service rates μ_1 and

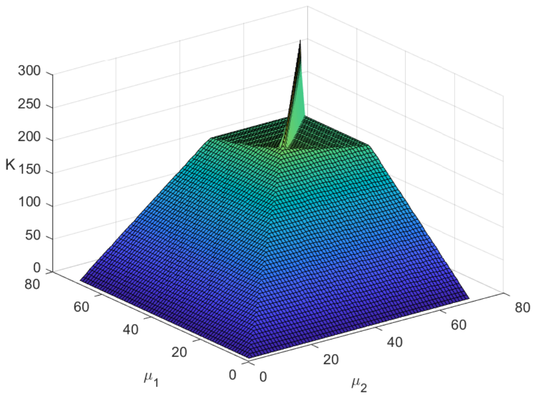
μ_2 vary in $\{1, 2, \dots, 70\}$, while the reward R is fixed, $R = 6$. With this setting, we study different situations: in the first case we suppose $C_1 < C_2$, for example $C_1 = 1$ and $C_2 = 2$, in the second case $C_1 > C_2$, for example $C_1 = 2$ and $C_2 = 1$, and in the third case $C_1 = C_2$, for example $C_1 = C_2 = 1.5$. The values of the equilibrium threshold in these three situations are represented in 3-dimensional plots (Figure 2.4). This numerical analysis was done with MATLAB (see [19]).



(a) For $C_1 = 1,\ C_2 = 2$.



(b) For $C_1 = 2,\ C_2 = 1$.



(c) For $C_1 = 1.5,\ C_2 = 1.5$.

Figure 2.4: Numerical computations of the equilibrium threshold K as μ_1 and μ_2 vary in $\{1, 2, \dots, 70\}$, with $R = 6$. Three different cases are analyzed and they are shown in (a), (b) and (c).

In all three cases, we observe that the behaviour for $\mu_1 = \mu_2$ is linear and it is determined by the equation $K = 4\mu_1 - 1$, which is coherent with Eq. (2.13). If instead, for $l = 1, 2$, we fix μ_l large and look at values of K as μ_{3-l} varies, we have that the behaviour still remains linear for μ_{3-l} small enough. Indeed, we expect it to be linear, because as μ_l goes to infinity, the model becomes an $M/M/1$ model with one queue with service rate μ_{3-l}, and in this model the behaviour of K is linear.

For example, in the first case we have that for μ_1 large, $K = 3\mu_2$, and for μ_2 large, $K = 6\mu_1 - 1$. In the second case, the behaviour is the opposite one, that is for μ_1 large, $K = 6\mu_2 - 1$, and for μ_2 large, $K = 3\mu_1$. Finally, in the third case the behaviour is similar in doing limits: for μ_1 large, $K = 4\mu_2$, and for μ_2 large, $K = 4\mu_1 - 1$.

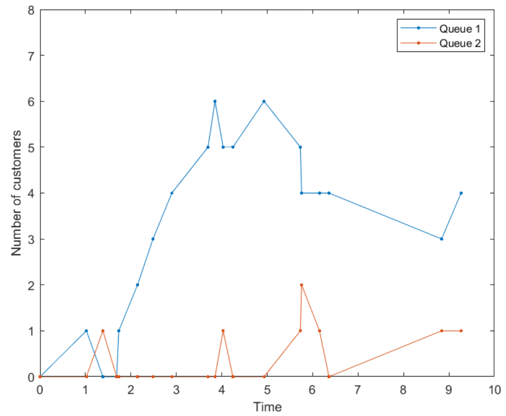
But we observe that, when μ_1 and μ_2 reach some values, K tends to stabilize at a certain level, and then slowly decreases. In our examples, except for $\mu_1 = \mu_2$, where the values of K continue to have a linear dependence, the equilibrium threshold K reaches a pick and then slowly decreases. In the first case and in the second one the pick is $K = 194$, while in the third one is $K = 183$.

Due to the symmetry of the expected sojourn times $T_1(k)$ and $T_2(k)$ with respect to μ_1 and μ_2, we have that the surfaces for the first two cases (Figure 2.4 (a) and Figure 2.4 (b)) are symmetric to each other with respect to the plane $\mu_1 = \mu_2$.

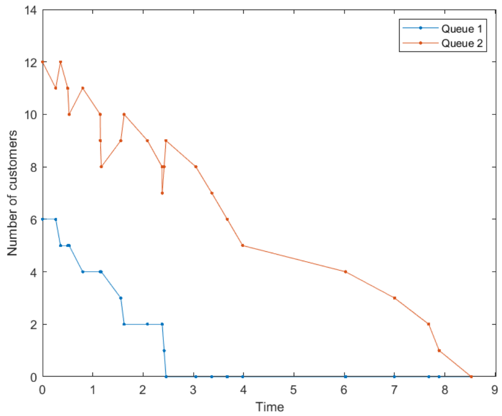
2.6 Numerical simulations

In this section, we compare theoretical results with results obtained from simulations. Indeed, we simulate the evolution of a two-node tandem network, which is initially empty and immediately filled by customers that arrive according to a Poisson process of rate $\lambda = 1$. The service rates are fixed and equal to $\mu_1 = 1$ and $\mu_2 = 2$. Assuming that customers are not using any strategy, that is they always join the system, Figure 2.5 (a) represents an example of the flow of customers in the network during a finite time period $T = 10$. With the same parameters, we simulate the route of a joining customer being in position $(6, 12)$ at her arrival until she leaves the system (Figure 2.5 (b)).

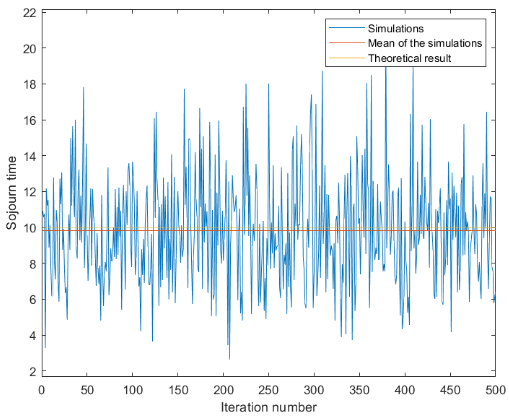
At this point, we compute also the time spent in the first queue and in the second queue by a joining customer, who observes the state $(6, 12)$ at her arrival. Indeed, in a simulation like the one in Figure 2.5 (b), the sojourn time in the first queue is given by the time taken for the blue line to became zero and, from that time until the red line becomes zero, the sojourn time in the second queue is computed. By doing multiple iterations, Figure 2.5 (c) and Figure 2.5 (d) represent the values of sojourn time in the first and second queue, respectively. Then we compare the average sojourn time coming from these iterations with the corresponding theoretical results (Lemma 2.1 and Lemma 2.3). We obtain that the simulations confirm the theoretical results. These plots are produced with MATLAB codes, which are available in GitHub [19].



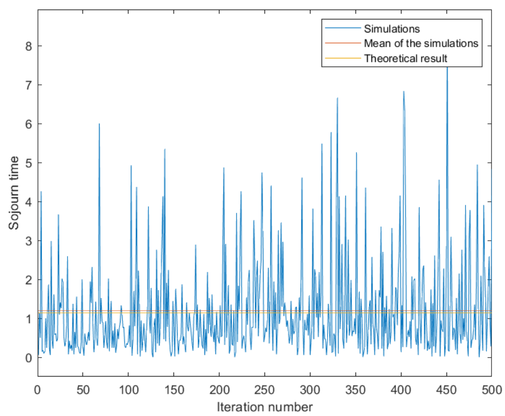
(a)



(b)



(c)



(d)

Figure 2.5: Considering a two-node tandem network with arrival rate $\lambda = 1$ and service rates $\mu_1 = 1$ and $\mu_2 = 2$, plot (a) simulates the evolution of the model which is initially empty and plot (b) the number of customer in front of the tagged one that joins the system in position $(6, 12)$. By repeating simulations like the one in plot (b), plot (c) and plot (d) show the sojourn times at queue 1 and at queue 2 and compare them with the corresponding theoretical results.

Chapter 3

Multi-node tandem network

Abstract

This chapter is a generalization of the previous one. We consider a multi-node tandem network, at which customers arrive according to a Poisson process. We prove that there exists an equilibrium strategy such that all customers can obtain an optimal profit based on a partial information, they receive upon arrival, about the system. The strategy indicates whether to join or not the system and we investigate if such strategy is a pure or mixed threshold one. The main reference is [12].

3.1 The model

After showing that in a two-node tandem network an equilibrium strategy exists under partial information about the state of the system, in this chapter we generalize this result to a multi-node tandem network.

A multi-node tandem network consists of M nodes, each corresponding to a queue with one server and an infinite capacity waiting line. We index the nodes by l, $l = 1, \dots, M$. The queues are arranged in tandem, that is, after a customer receives service at node l, she proceeds and joins queue $l+1$, if $l \leq M-1$; otherwise, she leaves the system. In each queue customers are served based on a FIFO discipline.

The service times of customers at each queue are assumed to be independent and exponentially distributed with mean μ_l^{-1}, that is the service rates are μ_l, $l = 1, \dots, M$. Customers arrive at the system according to a Poisson process with rate λ and, upon arrival, they receive partial information about the system, that is the total number of customers. Using this information, they decide whether to join the system or balk.

The customers' strategy for joining the system is represented by a sequence

$$s = (s_0, s_1, \dots), \tag{3.1}$$

where $s_k \in [0, 1]$ is the probability that an arriving customer joins the system, when the number of customers in the system is k.

The cost to a customer for staying at node l is C_l per unit of time. All customers have the same reward R from service completion. The profit of the tagged customer is given by

$$P = R - \sum_{l=1}^{M} C_l S_l,$$

where the random variable S_l denotes the sojourn time of the tagged customer at node l. The model is graphically represented in Figure 3.1.

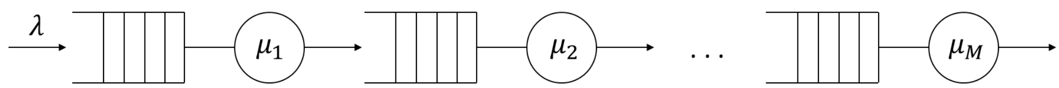


Figure 3.1: M-node tandem network with Poisson process of arrivals with rate λ and service times with rate μ_l, $l = 1, \dots, M$.

We look for a pure or mixed threshold strategy by maximizing the expected profit function of an arriving customer that observes k users in the system. The expected profit function under the strategy s is given by

$$P_s(k) = R - \sum_{l=1}^{M} C_l T_l(k),$$

where $T_l(k)$ is the expected sojourn time at queue l under the partial information. We assume that

$$R \geq \sum_{l=1}^{M} \frac{C_l}{\mu_l},$$

since, otherwise, even a customer that finds the system empty would get a negative profit by joining the system.

Remark 3.1 (K-threshold strategy). Using the notation in (3.1), we say that a strategy s is a pure K-threshold strategy if $s_i = 1$ for $i = 0, \dots, K-1$ and $s_i = 0$ otherwise, which means that all customers enter the system if and only if it contains less than K customers. Such a strategy is denoted by σ_K. Similarly, we say that a strategy s is a mixed x-threshold strategy, where $x = n + p$ with $n \in \mathbb{N}$ and $p \in (0,1)$, if $s_i = 1$ for $i = 0, \dots, n-1$, $s_n = p$, and $s_i = 0$ otherwise, which means that an arriving customer joins if there are less than n customers in the system, joins with probability p if there are exactly n customers, otherwise balks. Such a strategy is denoted by $\sigma_x = (1-p)\sigma_n + p\sigma_{n+1}$. So, a mixed threshold strategy is a convex combination of two pure threshold strategies.

3.2 Stationary distribution

Before making the decision to join the system or balk, customers receive the partial information about the state of the network, that is the total number of customers in the system. So, in this section, we are interested in studying the stationary distribution of customers in the system.

The state of the system at any time t can be described by a random vector $(Q_1(t), \ldots, Q_M(t))$, where $Q_l(t)$ represents the number of customers at node l at any time t. Under the strategy s, $\{(Q_1(t), \ldots, Q_M(t)) : t \geq 0\}$ is a continuous time Markov process, which is described by the following transition rates $q_s(n, n')$, with $n = (n_1, \ldots, n_M) \in \mathbb{Z}_+^M$ and $n' = (n'_1, \ldots, n'_M) \in \mathbb{Z}_+^M$

$$q_s(n,n') = \begin{cases} \lambda s_{|n|}, & \text{if } n' = n + e_1,\ |n| < K, \\ \mu_i, & \text{if } n' = n - e_i + e_{i+1},\ i = 1, \ldots, M-1, \\ \mu_M, & \text{if } n' = n - e_M, \\ -\lambda s_{|n|} - \sum_{i=1}^{M} \mu_i \mathbb{1}_{\{n_i > 0\}}, & \text{if } n' = n, \\ 0, & \text{otherwise}, \end{cases}$$

where $|n| = n_1 + \cdots + n_M$ and e_i is the M-dimensional unit vector with the i-th element equal to 1.

Suppose that the Markov process $\{(Q_1(t), \ldots, Q_M(t)) : t \geq 0\}$ has a stationary distribution. Let $\pi_s(n) = \mathbb{P}((Q_1^*, \ldots, Q_M^*) = n)$ be the stationary distribution, where Q_l^* is the number of customers at node l in the steady state. We set $Q^* = (Q_1^*, \ldots, Q_M^*)$ and $|Q^*| = Q_1^* + \cdots + Q_M^*$.

To give an explicit expression for the stationary distribution π_s, which depend on the strategy s, we first look at the stationary distribution π_{σ_K} in case all customers decide to follow a pure K-threshold strategy. In this situation, the network behaves like a semiopen Jackson network (Section 1.1), with an overall buffer capacity of K and routing probabilities $p_{ii+1} = 1$ for all $i = 0, \ldots, M-1$ and $p_{M0} = 1$. From the traffic equation

$$\lambda_i = \lambda p_{0i} + \sum_{j=1}^{M} \lambda_j p_{ji}, \quad i = 1, \ldots, M \tag{3.2}$$

we obtain that $\lambda_i = \lambda$ for $i = 1, \ldots, M$, and, applying Theorem 1.3, we get that

$$\pi_{\sigma_K}(n) = \frac{\prod_{l=1}^{M} \left(\frac{\lambda}{\mu_l}\right)^{n_l}}{\sum_{|n'| \leq K} \prod_{l=1}^{M} \left(\frac{\lambda}{\mu_l}\right)^{n'_l}}, \quad |n| \leq K.$$

Therefore, knowing this result, we state the following lemma, which gives an explicit formula for the stationary distribution, depending on the strategy s, adopted by all customers.

Lemma 3.2 (Stationary distribution)**.** If the tandem network has a stationary distribution under the strategy s, then the stationary distribution is given as follows: for $n \in \mathbb{Z}_+^M$

$$\pi_s(n) = \frac{\prod_{k=0}^{|n|} s_{k-1} \prod_{l=1}^{M} \left(\frac{\lambda}{\mu_l}\right)^{n_l}}{\sum_{n' \in \mathbb{Z}_+^M} \prod_{k=0}^{|n'|} s_{k-1} \prod_{l=1}^{M} \left(\frac{\lambda}{\mu_l}\right)^{n'_l}}, \tag{3.3}$$

where $s_{-1} = 1$ is fixed.

Proof: We prove formula (3.3) by using the definition: the stationary distribution $(\pi_s(n))_{n \in \mathbb{Z}_+^M}$ satisfies the *full balance equations*, which equates the "probability flow" out of each state n with the flow into the same state:

$$\pi_s(n) \sum_{n' \neq n} q_s(n, n') = \sum_{n' \neq n} \pi_s(n') q_s(n', n).$$

On the left hand side, we have

$$\pi_s(n) \left[\lambda s_{|n|} + \sum_{i=1}^{M-1} \mu_i \mathbb{1}_{\{n_i > 0\}} + \mu_M \mathbb{1}_{\{n_M > 0\}} \right],$$

while, on the right hand side, we have

$$\begin{aligned}
&\pi_s(n - e_1) q_s(n - e_1, n) \mathbb{1}_{\{n_1 > 0\}} + \sum_{i=1}^{M-1} \pi_s(n + e_i - e_{i+1}) q_s(n + e_i - e_{i+1}, n) \mathbb{1}_{\{n_{i+1} > 0\}} \\
&\quad + \pi_s(n + e_M) q_s(n + e_M, n) \\
&= c_s \Bigg[\prod_{k=0}^{|n|-1} s_{k-1} \prod_{l=1}^{M} \left(\frac{\lambda}{\mu_l}\right)^{(n-e_1)_l} \lambda s_{|n|-1} \mathbb{1}_{\{n_1 > 0\}} + \sum_{i=1}^{M-1} \prod_{k=0}^{|n|} s_{k-1} \prod_{l=1}^{M} \left(\frac{\lambda}{\mu_l}\right)^{(n+e_i-e_{i+1})_l} \mu_i \mathbb{1}_{\{n_{i+1} > 0\}} \\
&\quad + \prod_{k=0}^{|n|+1} s_{k-1} \prod_{l=1}^{M} \left(\frac{\lambda}{\mu_l}\right)^{(n+e_M)_l} \mu_M \Bigg] \\
&= \pi_s(n) \left[\lambda s_{|n|} + \sum_{i=2}^{M} \mu_i \mathbb{1}_{\{n_i > 0\}} + \mu_1 \mathbb{1}_{\{n_1 > 0\}} \right],
\end{aligned}$$

where c_s^{-1} is the normalization constant of Eq. (3.3)

$$c_s^{-1} = \sum_{n' \in \mathbb{Z}_+^M} \prod_{k=0}^{|n'|} s_{k-1} \prod_{l=1}^{M} \left(\frac{\lambda}{\mu_l}\right)^{n'_l}. \tag{3.4}$$

Since the two expressions are equal, $\pi_s(n)$ is the stationary distribution. □

Lemma 3.2 implies the following Corollary, where we add the partial information.

Corollary 3.3 (Stationary distribution under partial information)**.** If the tandem network has a stationary distribution under the strategy s, then the conditional joint distribution of Q^*, given $|Q^*| = k$, is

$$\pi(n \mid k) = \frac{\prod_{l=1}^{M} \left(\frac{1}{\mu_l}\right)^{n_l}}{\sum_{|n'|=k} \prod_{l=1}^{M} \left(\frac{1}{\mu_l}\right)^{n'_l}}, \quad |n| = k,$$

and, if moreover $s_i = 1$ for all $i = 0, 1, \dots, k-1$, the conditional joint distribution of Q^*, given $|Q^*| \leq k$, is

$$\pi^{\leq}(n \mid k) = \pi_{\sigma_k}(n).$$

Proof: Let c_s^{-1} be the normalization constant of (3.3), as in (3.4). Then, for $|n| = k$, we have that

$$\begin{aligned} \pi(n \mid k) &= \mathbb{P}(Q^* = n \mid |Q^*| = k) = \frac{\mathbb{P}(Q^* = n, |n| = k)}{\mathbb{P}(|Q^*| = k)} \\ &= \frac{c_s \prod_{j=0}^{k} s_{j-1} \prod_{l=1}^{M} \left(\frac{\lambda}{\mu_l}\right)^{n_l}}{\sum_{|n'|=k} c_s \prod_{j=0}^{k} s_{j-1} \prod_{l=1}^{M} \left(\frac{\lambda}{\mu_l}\right)^{n'_l}} = \frac{\prod_{l=1}^{M} \left(\frac{1}{\mu_l}\right)^{n_l}}{\sum_{|n'|=k} \prod_{l=1}^{M} \left(\frac{1}{\mu_l}\right)^{n'_l}}, \end{aligned} \tag{3.5}$$

and $\pi(n \mid k) = 0$, otherwise. While, for $|n| \leq k$, provided that $s_i = 1$ for all $i = 0, 1, \dots, k-1$, we have that

$$\begin{aligned} \pi^{\leq}(n \mid k) &= \mathbb{P}(Q^* = n \mid |Q^*| \leq k) = \frac{\mathbb{P}(Q^* = n, |n| \leq k)}{\mathbb{P}(|Q^*| \leq k)} \\ &= \frac{c_s \prod_{j=0}^{|n|} s_{j-1} \prod_{l=1}^{M} \left(\frac{\lambda}{\mu_l}\right)^{n_l}}{\sum_{|n'|\leq k} c_s \prod_{j=0}^{|n'|} s_{j-1} \prod_{l=1}^{M} \left(\frac{\lambda}{\mu_l}\right)^{n'_l}} = \frac{\prod_{l=1}^{M} \left(\frac{\lambda}{\mu_l}\right)^{n_l}}{\sum_{|n'|\leq k} \prod_{l=1}^{M} \left(\frac{\lambda}{\mu_l}\right)^{n'_l}} = \pi_{\sigma_k}(n), \end{aligned} \tag{3.6}$$

and $\pi^{\leq}(n \mid k) = 0$, otherwise. □

It follows that the distribution of $\pi(n \mid k)$ and $\pi^{\leq}(n \mid k)$ are independent of the strategy s and also of the arrival rate λ. This Corollary gives an important property which depend on the assumption of the model. For example, if the service times are Erlang, instead of exponential, we have that Corollary 3.3 doesn't hold (see Section 5.4).

3.3 Equilibrium strategies

We recall that Q^* denotes the stationary state of the system, which corresponds to the average state of the system at any instant. Not only, Q^* denotes also the average state of the system, observed by an arriving customer, since the arrival process is Poisson. This property is the so called *PASTA property* (Theorem 1.12). For this reason, we assume that, for an arriving user, the customer distribution in the system is the stationary one.

We suppose that all customers, besides the tagged one, follow a given strategy s. The tagged customer makes her decision based on the expected profit given the partial information she receives on the total number of customers in the network. The expected profit function $P_s(k)$, which may depend on the strategy s, is

$$P_s(k) = R - \sum_{l=1}^{M} C_l T_l(k) = R - C(k), \tag{3.7}$$

where $T_l(k) = \mathbb{E}_s[S_l \mid |Q^*| = k]$ is the expected sojourn time at queue l and $C(k) = \mathbb{E}_s[\sum_{l=1}^{M} C_l S_l \mid |Q^*| = k]$ is the expected cost for sojourn time in the system. Since we will see in Lemma 3.4 that $C(k)$ does not depend on the strategy s, the same holds for $P_s(k)$, thus, we will use $P(k)$ instead of $P_s(k)$. Of course, the strategy s must be compatible with the number of customers k. So, the strategy s must be such that $s_i > 0$ for all $i = 0, 1, \ldots, k-1$.

Let us suppose that all arriving customers use the strategy σ_{k+1}. Then $C^{\leq}(k) = \mathbb{E}_{\sigma_{k+1}}[\sum_{l=1}^{M} C_l S_l \mid |Q^*| \leq k]$ denotes the expected cost for sojourn time of a joining customer, that is of a customer that observes less than $k+1$ customers at her arrival. Let us first state and prove Lemma 3.4, and then, in Theorem 3.5, we give an explicit formula for computing $C(k)$.

Lemma 3.4 (Properties of $C(k)$ and $C^{\leq}(k)$)**.** The expected cost for sojourn time $C(k)$ is independent of the strategy s adopted by all customers. The expected cost for sojourn time for a customer that observes less than k customers at her arrival under the strategy σ_{k+1} is equal to the same one computed under the strategy σ_k, namely

$$\mathbb{E}_{\sigma_{k+1}}\left[\sum_{l=1}^{M} C_l S_l \mid |Q^*| \leq k-1\right] = \mathbb{E}_{\sigma_k}\left[\sum_{l=1}^{M} C_l S_l \mid |Q^*| \leq k-1\right] =: C^{\leq}(k-1).$$

Proof: We have that $C(k)$ does not depend on the strategy s adopted by all arriving customers, since

$$C(k) = \sum_{l=1}^{M} C_l T_l(k) = \sum_{l=1}^{M} C_l \sum_{|n|=k} \bar{T}_l(n + e_1)\pi(n \mid k),$$

where $\bar{T}_l(n + e_1)$ is the expected sojourn time of a joining customer observing the state $n \in \mathbb{Z}_+^M$ at her arrival and $\pi(n \mid k)$ is the conditional joint distribution defined in

Eq. (3.5), which does not depend on the strategy s, as proved in Corollary 3.3. So, by removing the subindex s, we can also write

$$C(k) = \mathbb{E}\Big[\sum_{l=1}^{M} C_l S_l \mid |Q^*| = k\Big].$$

For the second part of the statement, we have that the following expected value

$$\mathbb{E}_s\Big[\sum_{l=1}^{M} C_l S_l \mid |Q^*| \leq k-1\Big] = \sum_{j=0}^{k-1} C(j) \sum_{|n|=j} \pi^{\leq}(n \mid k-1)$$

does not depend on the strategy s, provided that $s_i = 1$ for all $i = 0, 1, \dots, k-1$, since both $C(j)$ and $\pi^{\leq}(n \mid k)$ doesn't depend on s, as proved in Corollary 3.3. So

$$\mathbb{E}_{\sigma_{k+1}}\Big[\sum_{l=1}^{M} C_l S_l \mid |Q^*| \leq k-1\Big] = \mathbb{E}_{\sigma_k}\Big[\sum_{l=1}^{M} C_l S_l \mid |Q^*| \leq k-1\Big] =: C^{\leq}(k-1).$$

□

Theorem 3.5 (Expected cost for sojourn time)**.** The expected cost for sojourn time of an arriving customer observing k customers in the system is

$$C(k) = \frac{\sum\limits_{|n|=k+1} \sum\limits_{i=1}^{M} n_i C_i \prod\limits_{l=1}^{M} \left(\frac{1}{\mu_l}\right)^{n_l}}{\sum\limits_{|n|=k} \prod\limits_{l=1}^{M} \left(\frac{1}{\mu_l}\right)^{n_l}}, \quad k = 0, 1, 2, \dots$$

Proof: We recall that, when all customers use the strategy s, from Lemma 3.4, $C(k)$ does not depend on s and, for $k = 0, 1, 2, \dots,$

$$C(k) = \sum_{l=1}^{M} C_l T_l(k) = \mathbb{E}\Big[\sum_{l=1}^{M} C_l S_l \mid |Q^*| = k\Big].$$

We will derive an explicit formula for $C(k)$. Let

$$x_j = \sum_{|n|=j} \prod_{l=1}^{M} \left(\frac{\lambda}{\mu_l}\right)^{n_l},$$

$$y_j = \sum_{|n|=j} \sum_{i=1}^{M} n_i C_i \prod_{l=1}^{M} \left(\frac{\lambda}{\mu_l}\right)^{n_l}.$$

As defined in *Remark* 3.1, let σ_{k+1} denote the threshold strategy under which customers join the network if and only if the total number of customers in the network is less than $k+1$, which means that $\sigma_{k+1} = (s_0, s_1, \dots)$ with $s_i = 1$ for $i = 0, 1, \dots, k$ and $s_i = 0$ otherwise. Let $C^{\leq}(k)$ be the expected cost for sojourn time incurred by a joining customer under the threshold strategy σ_{k+1}.

Under the threshold strategy σ_{k+1}, the probability of the total number of customers in the network being i, for $i = 0, \dots, k+1$, is

$$\mathbb{P}_{\sigma_{k+1}}(|Q^*| = i) = \sum_{|n|=i} \pi_{\sigma_{k+1}}(n) = \frac{\sum\limits_{|n|=i} \prod\limits_{l=1}^{M} \left(\frac{\lambda}{\mu_l}\right)^{n_l}}{\sum\limits_{j=0}^{k+1} \sum\limits_{|n|=j} \prod\limits_{l=1}^{M} \left(\frac{\lambda}{\mu_l}\right)^{n_l}} = \frac{x_i}{\sum\limits_{j=0}^{k+1} x_j},$$

and 0 if $i > k+1$, where we have used formula (3.3) with $s = \sigma_{k+1}$. Hence, the joining probability of an arbitrary arriving customer, under the threshold strategy σ_{k+1}, is

$$\mathbb{P}_{\sigma_{k+1}}(|Q^*| \leq k) = \sum_{j=0}^{k} \mathbb{P}_{\sigma_{k+1}}(|Q^*| = j) = \frac{\sum\limits_{j=0}^{k} x_j}{\sum\limits_{j=0}^{k+1} x_j}.$$

and so the joining rate of customers $\bar{\lambda}$ is

$$\bar{\lambda} = \lambda \frac{\sum_{j=0}^{k} x_j}{\sum_{j=0}^{k+1} x_j} = \lambda\left(1 - \frac{x_{k+1}}{\sum_{j=0}^{k+1} x_j}\right).$$

The expectation of $\sum_{i=1}^{M} C_i Q_i^*$, under the threshold strategy σ_{k+1}, is

$$\mathbb{E}_{\sigma_{k+1}}\left[\sum_{i=1}^{M} C_i Q_i^*\right] = \sum_{n \in \mathbb{Z}_+^M} \sum_{i=1}^{M} C_i \; n_i \; \pi_{\sigma_k}(n) = \frac{\sum\limits_{j=0}^{k+1} \sum\limits_{|n|=j} \sum\limits_{i=1}^{M} n_i C_i \prod\limits_{l=1}^{M} \left(\frac{\lambda}{\mu_l}\right)^{n_l}}{\sum\limits_{j=0}^{k+1} \sum\limits_{|n|=j} \prod\limits_{l=1}^{M} \left(\frac{\lambda}{\mu_l}\right)^{n_l}} = \frac{\sum\limits_{j=0}^{k+1} y_j}{\sum\limits_{j=0}^{k+1} x_j}.$$

From Little's law, as defined in Eq. (1.8), with $L = \mathbb{E}_{\sigma_{k+1}}[\sum_{i=1}^{M} C_i Q_i^*]$, $T = C^{\leq}(k)$ and average arrival rate $\bar{\lambda}$, we have that

$$C^{\leq}(k) = \frac{\mathbb{E}_{\sigma_{k+1}}\left[\sum_{i=1}^{M} C_i Q_i^*\right]}{\bar{\lambda}} = \frac{\sum_{j=0}^{k+1} y_j}{\lambda \sum_{j=0}^{k} x_j}. \tag{3.8}$$

Under the threshold strategy σ_{k+1}, $k \geq 1$, the probability that a joining customer has the information that there are less than k customers in the network is

$$\mathbb{P}_{\sigma_{k+1}}(|Q^*| < k \mid |Q^*| \leq k) = \frac{\mathbb{P}(|Q^*| < k)}{\mathbb{P}(|Q^*| \leq k)} = \frac{\sum_{j=0}^{k-1} x_j}{\sum_{j=0}^{k+1} x_j} \cdot \frac{\sum_{j=0}^{k+1} x_j}{\sum_{j=0}^{k} x_j} = \frac{\sum_{j=0}^{k-1} x_j}{\sum_{j=0}^{k} x_j}.$$

and the probability that a joining customer has the information that there are exactly k customers in the network is

$$\mathbb{P}_{\sigma_{k+1}}(|Q^*| = k \mid |Q^*| \leq k) = \frac{\mathbb{P}(|Q^*| = k)}{\mathbb{P}(|Q^*| \leq k)} = \frac{x_k}{\sum_{j=0}^{k+1} x_j} \cdot \frac{\sum_{j=0}^{k+1} x_j}{\sum_{j=0}^{k} x_j} = \frac{x_k}{\sum_{j=0}^{k} x_j}.$$

Therefore, we have that, for $k = 1, 2, \ldots$

$$C^{\leq}(k) = \frac{\sum_{j=0}^{k-1} x_j}{\sum_{j=0}^{k} x_j} \mathbb{E}_{\sigma_{k+1}}\left[\sum_{l=1}^{M} C_l S_l \mid |Q^*| \leq k-1\right] + \frac{x_k}{\sum_{j=0}^{k} x_j} \mathbb{E}_{\sigma_{k+1}}\left[\sum_{l=1}^{M} C_l S_l \mid |Q^*| = k\right],$$

which, from Lemma 3.4, is equivalent to

$$C^{\leq}(k) = \frac{\sum_{j=0}^{k-1} x_j}{\sum_{j=0}^{k} x_j} C^{\leq}(k-1) + \frac{x_k}{\sum_{j=0}^{k} x_j} C(k). \tag{3.9}$$

Substituting the expression (3.8) into Eq. (3.9), we obtain that

$$C(k) = \frac{y_{k+1}}{\lambda x_k} = \frac{\sum\limits_{|n|=k+1} \sum\limits_{i=1}^{M} n_i C_i \prod\limits_{l=1}^{M} \left(\frac{1}{\mu_l}\right)^{n_l}}{\sum\limits_{|n|=k} \prod\limits_{l=1}^{M} \left(\frac{1}{\mu_l}\right)^{n_l}}, \quad k = 1, 2, \ldots,$$

and it also holds for $k = 0$, because for $k = 0$ we have that

$$C(0) = \sum_{i=1}^{M} \frac{C_i}{\mu_i},$$

and the proof is complete. □

From Theorem 3.5 and Eq. (3.7), we obtain that $P(k)$ can be written as

$$P(k) = R - \frac{\sum\limits_{|n|=k+1} \sum\limits_{i=1}^{M} n_i C_i \prod\limits_{l=1}^{M} \left(\frac{1}{\mu_l}\right)^{n_l}}{\sum\limits_{|n|=k} \prod\limits_{l=1}^{M} \left(\frac{1}{\mu_l}\right)^{n_l}}, \quad k = 0, 1, 2, \ldots \tag{3.10}$$

We observe that $P(k)$ is strictly decreasing and goes to $-\infty$ as $k \to +\infty$. A formal proof of these properties is given in Theorem 4.5. Finally, we state the main result that says which are the equilibrium strategies.

Theorem 3.6 (Equilibrium strategies)**.** Let K be the threshold given by

$$K = \inf\{k \in \mathbb{Z}_+ : P(k) < 0\}. \tag{3.11}$$

If $P(K-1) > 0$, then the threshold strategy σ_K is the only equilibrium strategy. If $P(K-1) = 0$, then all the threshold strategies $\sigma_{(K-1)+p} = (1-p)\sigma_{K-1} + p\sigma_K$, with $p \in [0, 1]$, are the only equilibrium strategies.

Proof: Let $F_{s^1,s^2}(k)$ be the payoff for a customer that chooses to follow the strategy s^1, when everyone else chooses to follow the strategy s^2. In this case, the payoff function corresponds to $P(k)$ if the strategy s^1 says to enter the system, 0 otherwise, that is

$$F_{s^1,s^2}(k) = P(k)X(k),$$

where $X(k)$ is a Bernoulli random variable of parameter s^1_k.

So, from Definition 1.17, we have that a strategy s is an *equilibrium strategy* if and only if

$$\begin{aligned} s_k &= 1 \quad \text{for } k \text{ with } P(k) > 0 \quad \text{and} \\ s_k &= 0 \quad \text{for } k \text{ with } P(k) < 0, \end{aligned} \tag{3.12}$$

and the proof is completed. □

So, when $P(K-1) > 0$, there is a unique equilibrium strategy, which is a pure threshold strategy, while in the second case, that is when $P(K-1) = 0$, there are two pure threshold strategies and also mixed threshold strategies, which are given by any strictly convex combination of the two pure threshold strategies.

Furthermore, these strategies are also *dominant strategies.*

We analyze a particular example of multi-node tandem network. We assume that all service rates and all costs for sojourn time are equal to each other, that is $\mu_i = \mu_j =: \mu$ and $C_i = C_j =: C$ for all $i, j = 1, \ldots, M$. Then, from Eq. (3.10), the expected profit function $P(k)$ becomes

$$P(k) = R - \frac{(k+1)C}{\mu} \cdot \frac{\sum\limits_{|n|=k+1} 1}{\sum\limits_{|n|=k} 1} = R - \frac{C}{\mu}(k+M),$$

where we use *Remark* 3.7. In this case the equilibrium threshold defined in (3.11) is

$$K = \left\lfloor \frac{R\mu}{C} \right\rfloor - M + 1.$$

We can compare this result with the one found in the observable $M/M/1$ model, where the equilibrium threshold is (1.6), and with the two-node tandem network, where the equilibrium threshold is (2.14). We can therefore observe that, if we consider tandem networks where queues have the same service rates and the same costs for sojourn time, then, as the number of tandem nodes grows, the optimal threshold for entering the system decreases linearly.

Remark 3.7 (How many M-dimensional vectors sum to k?). The number of M-dimensional vectors with integer entries greater than or equal to 0, whose sum is equal to k, is

$$\sum_{|n|=k} 1 = \binom{k+M-1}{k}. \tag{3.13}$$

Formula (3.13) can be proved by an induction argument.

Now, we analyze the situation for μ_j goes to infinity, for $j = 1, \dots, M$. We expect that the model behaves as a model with $M-1$ queues, that is as a multi-node tandem network without the j-th queue. Indeed, as soon as a customer reaches queue j, she immediately moves to queue $j+1$, if $j < M$, or leaves the system, if $j = M$, so queue j does not affect the sojourn of a customer in the system. To prove this result, we compute the limit of $C(k)$ as $\mu_j \to +\infty$. We obtain that

$$\lim_{\mu_j \to +\infty} C(k) = \frac{\sum\limits_{|n|=k+1} \sum\limits_{i=1, i\neq j}^{M} n_i C_i \prod\limits_{l=1}^{M} \left(\frac{1}{\mu_l}\right)^{n_l}}{\sum\limits_{|n|=k} \prod\limits_{l=1}^{M} \left(\frac{1}{\mu_l}\right)^{n_l}}, \quad k = 0, 1, 2, \dots, \tag{3.14}$$

where the sum over $|n| = k+1$ and $|n| = k$ is made for $n \in \mathbb{Z}_+^{M-1}$. That is the limit in Eq. (3.14) corresponds exactly to the expected cost for sojourn time in a multi-node tandem network with $M-1$ queues obtained by removing queue j.

Multi-node tandem with exits

We give the same results for a multi-node tandem network with exits. For the proof, see Chapter 4, where a more extended model is analyzed.

We consider a multi-node tandem network, with the same structure and the same notations used in Section 3.1, but we suppose that, after the completion of the service at queue l, a customer joins the next queue $l+1$ with probability p_l and leaves the system with probability $\bar{p}_l = 1 - p_l$, if $l \leq M-1$; while for $l = M$, the customer can only leaves the system.

The model is represented in the Figure 3.2, where customers at each queue can leave the system after the completion of the service.

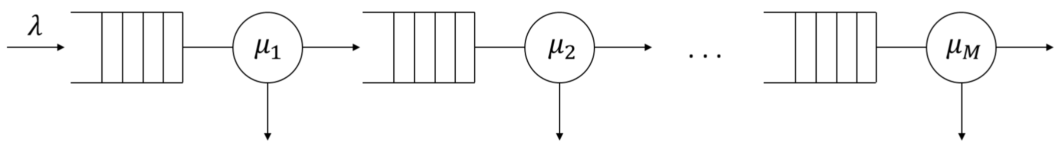


Figure 3.2: M-node tandem network with exits. The arrival process is a Poisson process with rate λ and the service times are exponential variables with rate μ_l.

Then, the expected profit $P(k)$ of a joining customer who observes k customers in the system at her arrival is

$$P(k) = R - \frac{\sum\limits_{|n|=k+1} \sum\limits_{i=1}^{M} n_i C_i \prod\limits_{l=1}^{M} \prod\limits_{j=0}^{l-1} \left(\frac{p_j}{\mu_l}\right)^{n_l}}{\sum\limits_{|n|=k} \prod\limits_{l=1}^{M} \prod\limits_{j=0}^{l-1} \left(\frac{p_j}{\mu_l}\right)^{n_l}}. \tag{3.15}$$

where $p_0 = 1$, since an arriving customer is definitely routed to node 1 if she decides to join the system. Also for this model, Theorem 3.6 holds with the expected profit function defined in Eq. (3.15).

3.4 Numerical computations

In this section, we analyze some particular situations. First, we fix all the parameters, except the number of nodes M, and, then, we look at how the equilibrium threshold changes as M increases. Secondly, we consider a tandem network with three nodes, fix the parameters of reward and costs for sojourn time and look at the behaviour of the equilibrium threshold, as the service rates changes.

For the first situation, we take $R = 3M$, and, for $l = 1, \ldots, M$, $\mu_l = l$ and $C_l = 1$. The following Table 3.1 lists the equilibrium thresholds K for different values of M. We observe that the equilibrium threshold K grows almost linearly, as M varies from 1 to 10.

M	1	2	3	4	5	6	7	8	9	10
K	3	5	8	11	14	17	20	23	26	29

Table 3.1: The values of equilibrium thresholds K for different values of M.

For the second situation, we fix $M = 3$, $R = 9$ and $C_l = 1$ for $l = 1, 2, 3$. Figure 3.3 represents the thresholds of the equilibrium strategy for different values of μ_1 and μ_2, as μ_3 vary. We observe that for every choice of μ_1 and μ_2 the behaviour of K as a function of μ_3 is linear for some values and, then, at a certain level it stabilizes, but as μ_1 and μ_2 increase, the level, in which the equilibrium threshold stabilizes, grows. In particular, if μ_1 and μ_2 are fixed, we have that the values of the equilibrium threshold start to stabilize when $\mu_3 = \min\{\mu_1, \mu_2\}$. All the MATLAB codes that produce these plots are available in GitHub [19].

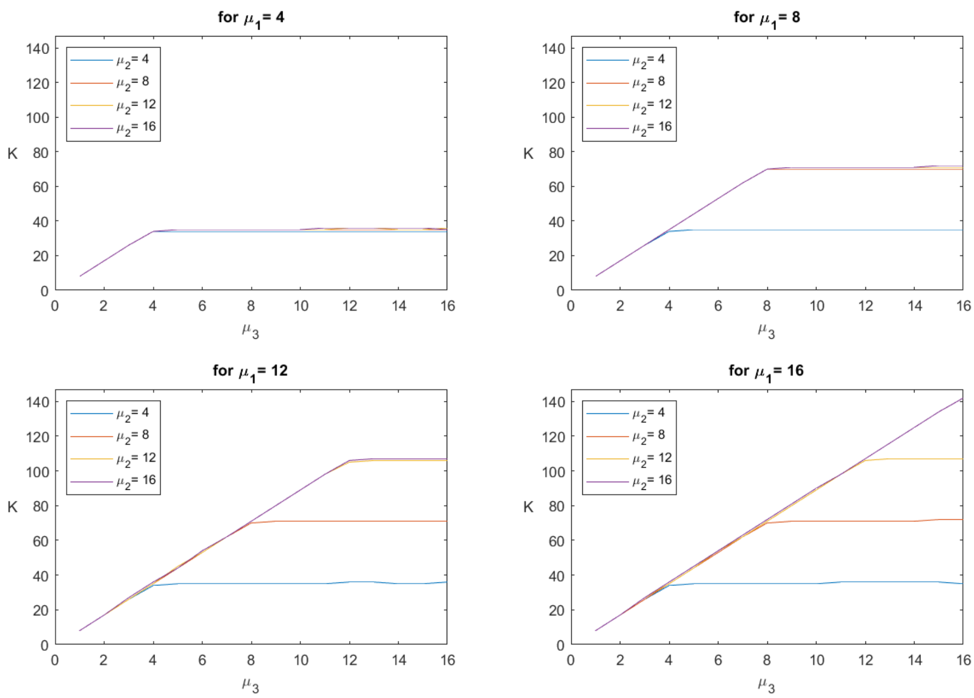


Figure 3.3: Numerical computations of the equilibrium threshold K as μ_1 and μ_2 vary in $\{4, 8, 12, 16\}$, and μ_3 in $\{1, 2, \ldots, 16\}$ with $R = 9$.

Chapter 4

Tree network

Abstract

We consider a tree network and compute the equilibrium strategies for customers who all use the same strategy, by maximizing the expected profit function. We show that a pure or mixed threshold strategy exists when customers only have partial information about the number of users in the system upon their arrival. We also compare this model with a tandem network.

4.1 The model

In Chapters 2 and 3, the equilibrium strategies are studied for tandem network [7, 12]. In this chapter we analyze more general overtaking free queueing networks, which are characterized by a branching structure, the so called *tree networks*. Tree networks are particular queueing networks, where different paths are possible for a customer to complete services and exit the system. In particular, in a tree network we assume that the structure of the queueing network is an *out-tree*, as defined in Section 1.7.

It means that, differently from a tandem network where all joining customers follow the same path in the system, in a tree network joining customers can move along the nodes with different paths. This model can be viewed as an extension of the multi-node tandem network with exits. As a matter of fact, a customer that is routed along a specific path, looks at others that follow different paths as if they exit the system. This extension is possible because customers in front of the tagged one will remain always in front of her, if they follow the same path of the tagged one, otherwise, when they come out from the path of the tagged customer, they will never appear in front of her again. Regarding those customers behind the tagged one, they are not of interest in order to compute the expected sojourn time at any queue.

Indeed, the aim of this chapter is to compute the equilibrium threshold strategy according to which a customer can decide whether to join or balk the system only knowing the total number of customers in the system. It means that an arriving customer is informed about the number of customers in the system and she immediately decides whether to join or not, only knowing this partial information. If she enters the system, she can not renege to her decision, if she doesn't, she can't get back to the system ever again. The optimal strategy is computed assuming that all customers follow the same strategy. So, we build an expected profit function P, which depends only on the number of users in the system upon a customer's arrival, and find the threshold strategy that gives the equilibrium.

Let us consider a tree network and denote with $G = (V, E)$ the associated out-tree, whose vertices correspond to the nodes where customers can queue up and receive the service, and whose edges correspond to all possible movements from one queue to another. We denote with ν and ε the cardinality of V and E, respectively. All customers, if they enter the system, join the queue at the first node, that we call node 1, and then they will be routed to other nodes, following a particular path. We recall that Lemma 1.29 implies that there is only one path that connects node 1 to any other node.

We use $S^-(i) = \{j \in V : (i, j) \in E\}$ to denote the set of vertices reachable from a vertex $i \in V$, that is, after completing the service at queue i, a customer either joins a queue among nodes in $S^-(i)$ or leaves the system; and $d^-(i) = |S^-(i)|$ denotes the *outdegree* of vertex i, that is the number of reachable vertices from vertex i. We also denote with $V_2 \subset V$ the set of leafs, that is $V_2 = \{i \in V : d^-(i) = 0\}$; and with $V_1 = V \setminus V_2$ the rest of the vertices.

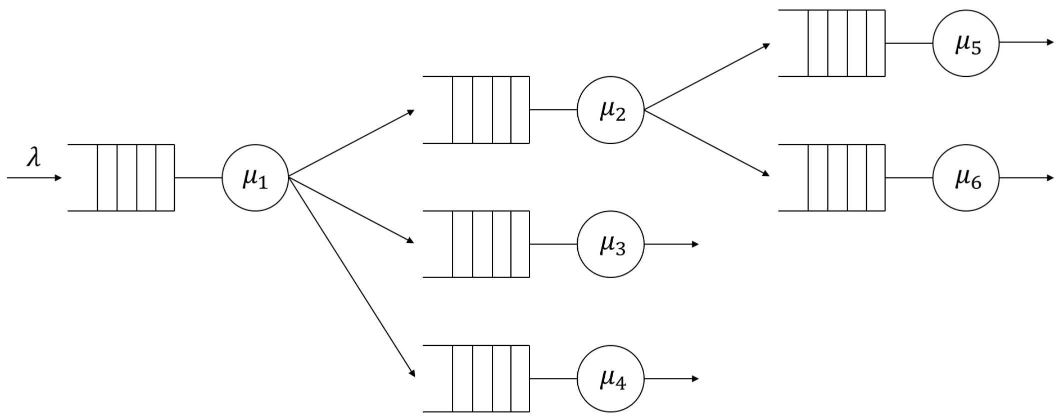
In this model, customers arrive to the system according to a Poisson process with intensity λ and are served based on a FIFO discipline. If they join the system, they immediately log into the queue of node 1 and, after the completion of the service at each node $i \in V$, they are routed to another queue at node $j \in S^-(i)$ with probability p_{ij}, otherwise they leave the system with probability $p_{i0} = 1 - \sum_{j \in S^-(i)} p_{ij}$. We also put $p_{01} = 1$ and $p_{i0} = 1$ for all $i \in V_2$. Node 0 represents the outside environment, from which customers arrive at the system and into which customers go out. Probabilities $(p_{ij})_{i \in V_1, j \in S^-(i)}$ are the routing probabilities and they describes the flow of customers in the system. The service times of customers are independent and exponentially distributed with mean μ_i^{-1}, for $i \in V$.

An example of this situation is in Figure 4.1 (a), where we choose the graph $G = (V, E)$, with $V = \{1, 2, 3, 4, 5, 6\}$ and $E = \{(1, 2), (1, 3), (1, 4), (2, 5), (2, 6)\}$ as the out-tree describing the tree network. Considering the same example, in Figure 4.1 (b), the corresponding routing network is graphically represented.

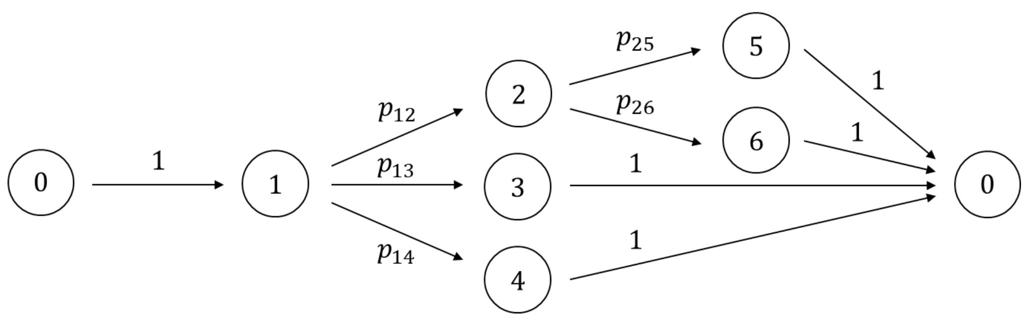
4.2 Stationary distribution

Let $s = (s_0, s_1, \dots)$ be a strategy, where s_k corresponds to the probability of a customer to join the system, knowing that the total number of customers in the system is k. We recall that σ_K denotes the K-threshold strategy, as defined in Remark 3.1.

Let $P_s(k)$ be the expected profit function for a joining customer, that upon her



(a) The queueing network.



(b) The routing network.

Figure 4.1: An example of tree network, where the underlying out-tree is the graph $G = (V, E)$, with $V = \{1, 2, 3, 4, 5, 6\}$ and $E = \{(1, 2), (1, 3), (1, 4), (2, 5), (2, 6)\}$. Figure (a) represents the structure of the queueing network, while figure (b) represents the routing network, where customers move from node i to node j with probability p_{ij}. Node 0 represents the outside environment.

arrival observes k users in the network, namely

$$P_s(k) = R - C(k) = R - \sum_{i \in V} C_i T_i(k), \tag{4.1}$$

where R is the reward a customer gets if she enters the system, C_i is the unitary cost for sojourn time at queue i, $T_i(k)$ is the expected sojourn time at queue i, knowing the partial information, and $C(k) = \sum_{i \in V} C_i T_i(k)$ is the expected cost for sojourn time. We will see that $P_s(k)$ does not depend on the strategy s, so we denote it with $P(k)$.

We compute an explicit expression for $P(k)$ and the equilibrium strategy is found by maximizing the expected profit function, assuming that all customers use the same strategy. We also suppose that a customer, that doesn't enter the system, gets a profit equal to 0.

The state of the system at any time t can be described by a random vector $(Q_i(t))_{i\in V}$, where $Q_i(t)$ is the number of customers at node i, at time t. Under the strategy s, $\{(Q_i(t))_{i\in V} : t \geq 0\}$ is a continuous time Markov process, which is described by the transition rates $q_s(n, n')$, with $n, n' \in \mathbb{Z}_+^\nu$. The transition rates are

$$q_s(n,n') = \begin{cases} \lambda s_{|n|}, & \text{if } n' = n + e_1 \\ p_{ij}\mu_i, & \text{if } n' = n - e_i + e_j,\ i \in V_1,\ j \in S^-(i) \\ p_{i0}\mu_i, & \text{if } n' = n - e_i,\ i \in V \\ -\lambda s_{|n|} - \sum_{i\in V} \mu_i \mathbb{1}_{\{n_i>0\}}, & \text{if } n' = n \\ 0, & \text{otherwise,} \end{cases}$$

where $|n| = \sum_{i\in V} n_i$ and e_i is the ν-dimensional unit vector with the i-th element equal to 1. We observe that the first transition rate refers to the arrival rate at the system, the second one refers to the transition rate for moving from a queue to another and the third one corresponds to the transition rate for exiting the system. Regarding the fourth equation, we observe that

$$\sum_{i\in V} \mu_i \mathbb{1}_{\{n_i>0\}} = \sum_{i\in V_1} \sum_{j\in S^-(i)} p_{ij}\mu_i \mathbb{1}_{\{n_i>0\}} + \sum_{i\in V} p_{i0}\mu_i \mathbb{1}_{\{n_i>0\}}.$$

Indeed

$$\begin{aligned} &\sum_{i\in V_1} \sum_{j\in S^-(i)} p_{ij}\mu_i \mathbb{1}_{\{n_i>0\}} + \sum_{i\in V} p_{i0}\mu_i \mathbb{1}_{\{n_i>0\}} \\ =&\sum_{i\in V_1} \sum_{j\in S^-(i)} p_{ij}\mu_i \mathbb{1}_{\{n_i>0\}} + \sum_{i\in V_1} p_{i0}\mu_i \mathbb{1}_{\{n_i>0\}} + \sum_{i\in V_2} p_{i0}\mu_i \mathbb{1}_{\{n_i>0\}} \\ =&\sum_{i\in V_1} \Big(\sum_{j\in S^-(i)} p_{ij} + p_{i0} \Big) \mu_i \mathbb{1}_{\{n_i>0\}} + \sum_{i\in V_2} \mu_i \mathbb{1}_{\{n_i>0\}} \\ =&\sum_{i\in V_1} \mu_i \mathbb{1}_{\{n_i>0\}} + \sum_{i\in V_2} \mu_i \mathbb{1}_{\{n_i>0\}} = \sum_{i\in V} \mu_i \mathbb{1}_{\{n_i>0\}}, \end{aligned}$$

where we use the fact that $\sum_{j\in S^-(i)} p_{ij} + p_{i0} = 1$, if $i \in V_1$, and $p_{i0} = 1$, otherwise.

Suppose that the Markov process $\{(Q_i(t))_{i\in V} : t \geq 0\}$ has a stationary distribution. Let $\pi_s = (\pi_s(n))_{n\in\mathbb{Z}_+^\nu}$ be the stationary distribution, that is, for $n \in \mathbb{Z}_+^\nu$, $\pi_s(n) = \mathbb{P}(Q^* = n)$ is the stationary probability of observing the state n in the system, where $Q^* = (Q_i^*)_{i\in V}$ is the vector with Q_i^* being the number of customers at node i in the steady state. We also set $|Q^*| = \sum_{i\in V} Q_i^*$.

To compute the stationary distribution, we first calculate the arrival rate to node $i \in V$, called λ_i, which is obtained by solving the traffic equation (see Eq. 1.1)

$$\lambda_i = \lambda p_{0i} + \sum_{j\in V} \lambda_j p_{ji},$$

which is equivalent to

$$\begin{cases} \lambda_1 = \lambda \\ \lambda_j = \lambda_i p_{ij}, \quad \text{for all } i \in V_1,\ j \in S^-(i) \end{cases} \tag{4.2}$$

which means that, if $(i,j) \in E$ is an edge of the underlying out-tree of the tree network, then $\lambda_j = \lambda_i p_{ij}$, that is from node i to node j the arrival rate is rescaled by a factor of p_{ij}.

Now, we denote with $W(i)$ the sequence of vertices of V that form the path from node 1 to node i, that is

$$W(i) = \{1 = v_0(i), v_1(i), \ldots, v_n(i) = i : (v_{j-1}(i), v_j(i)) \in E \forall j = 1, \ldots, n\}.$$

Then, for all $i \in V \setminus \{1\}$,

$$\lambda_i = \prod_{k=1}^{|W(i)|-1} p_{v_{k-1}(i)v_k(i)} \lambda, \quad \text{and} \quad \lambda_1 = \lambda.$$

Let us call $q_1 := 1$ and $q_i := \prod_{k=1}^{|W(i)|-1} p_{v_{k-1}(i)v_k(i)}$ for $i \in V \setminus \{1\}$. We have that, for $i \in V$, q_i corresponds to the probability of reaching node i from node 0. In particular, for $i \in V \setminus \{1\}$, q_i is the probability that a joining customer follows the path from node 1 to node i. As a consequence of this setting, it turns out that for all $i \in V$

$$\lambda_i = q_i \lambda.$$

By the definition of $(q_i)_{i\in V}$, we have that

$$\sum_{i\in V} q_i p_{i0} = 1, \tag{4.3}$$

because $q_i p_{i0}$ is the probability of exiting the system at node i and, so, since any customer leaves the system in a finite time period, the sum is equal to 1.

If we consider the tree network as a semiopen Jackson network (Section 1.1), with an overall buffer capacity of K, which means that the adopted strategy by all customers is the strategy σ_K, then we get that, from Theorem 1.3, the stationary distribution is

$$\pi_{\sigma_K}(n) = \frac{\prod_{i\in V} \left(\frac{\lambda_i}{\mu_i}\right)^{n_i}}{\sum_{|n'|\le K} \prod_{i\in V} \left(\frac{\lambda_i}{\mu_i}\right)^{n'_i}}, \quad |n| \le K. \tag{4.4}$$

But in general, the stationary distribution depends on the strategy s and it is given by the following theorem.

Lemma 4.1 (Stationary distribution)**.** If the tree network has a stationary distribution under the strategy s, then the stationary distribution is

$$\pi_s(n) = \frac{\prod_{k=0}^{|n|} s_{k-1} \prod_{i\in V} \left(\frac{\lambda_i}{\mu_i}\right)^{n_i}}{\sum_{n'\in\mathbb{Z}_+^M} \prod_{k=0}^{|n'|} s_{k-1} \prod_{i\in V} \left(\frac{\lambda_i}{\mu_i}\right)^{n'_i}}, \quad n \in \mathbb{Z}_+^\nu, \tag{4.5}$$

where $s_{-1} = 1$ is fixed.

Proof: To prove formula (4.5), we use the definition: the stationary distribution π_s satisfies the *full balance equations*, which equates the "probability flow" out of each state n with the flow into the same state:

$$\pi_s(n) \sum_{n'\neq n} q_s(n, n') = \sum_{n'\neq n} \pi_s(n') q_s(n', n). \tag{4.6}$$

So we put (4.5) into Eq. (4.6) and verify that the equality holds. On the left hand side we have

$$\pi_s(n) \left[\lambda s_{|n|} + \sum_{i\in V} \mu_i \mathbb{1}_{\{n_i>0\}} \right],$$

while on the right hand side, we have

$$\begin{aligned}
&\pi_s(n-e_1)q_s(n-e_1,n)\mathbb{1}_{\{n_1>0\}} + \sum_{i\in V_1}\sum_{j\in S^-(i)} \pi_s(n+e_i-e_j)q_s(n+e_i-e_j,n)\mathbb{1}_{\{n_j>0\}} \\
&+ \sum_{i\in V} \pi_s(n+e_i)q_s(n+e_i,n) \\
=&c_s\Bigg[\prod_{k=0}^{|n|-1} s_{k-1} \prod_{\ell\in V} \left(\frac{\lambda_\ell}{\mu_\ell}\right)^{(n-e_1)_\ell} \lambda s_{|n|-1}\mathbb{1}_{\{n_1>0\}} + \sum_{i\in V_1}\sum_{j\in S^-(i)} \prod_{k=0}^{|n|} s_{k-1} \prod_{\ell\in V} \left(\frac{\lambda_\ell}{\mu_\ell}\right)^{(n+e_i-e_j)_\ell} p_{ij}\mu_i \mathbb{1}_{\{n_j>0\}} \\
&+ \sum_{i\in V} \prod_{k=0}^{|n|+1} s_{k-1} \prod_{\ell\in V} \left(\frac{\lambda_\ell}{\mu_\ell}\right)^{(n+e_i)_\ell} p_{i0}\mu_i \Bigg] \\
=&\pi_s(n)\left[s_{|n|-1}^{-1} \frac{\mu_1}{\lambda_1} \lambda s_{|n|-1}\mathbb{1}_{\{n_1>0\}} + \sum_{i\in V_1}\sum_{j\in S^-(i)} \frac{\lambda_i}{\mu_i}\frac{\mu_j}{\lambda_j} p_{ij}\mu_i\mathbb{1}_{\{n_j>0\}} + \sum_{i\in V} s_{|n|}\frac{\lambda_i}{\mu_i} p_{i0}\mu_i \right] \\
=&\pi_s(n)\left[\lambda s_{|n|} + \sum_{i\in V} \mu_i \mathbb{1}_{\{n_i>0\}} \right],
\end{aligned}$$

where, in the third step, we use the relations (4.2) and (4.3), while c_s^{-1} is the denominator of Eq. (4.5). Since the two expressions are equal, $\pi_s(n)$ is the stationary distribution we were looking for. □

We observe that the distribution (4.5) with the strategy $s = \sigma_K$ gives the formula in (4.4). If, instead, we compute the stationary distribution conditioned on the total number of customers in the system, say k, then the stationary distribution does not depend on the strategy s and its expression is given by the following theorem.

Corollary 4.2 (Stationary distribution under partial information)**.** If the tree network has a stationary distribution under the strategy s, then the conditional joint distribution of Q^*, given $|Q^*| = k$, is

$$\pi(n \mid k) = \frac{\prod_{i=1}^{M} \left(\frac{q_i}{\mu_i}\right)^{n_i}}{\sum_{|n'|=k} \prod_{i=1}^{M} \left(\frac{q_i}{\mu_i}\right)^{n'_i}}, \quad |n| = k, \ n \in \mathbb{Z}_+^{\nu},$$

which does not depend on the strategy s. If moreover $s_i = 1$ for all $i = 0, 1, \ldots, k-1$, the conditional joint distribution of Q^*, given $|Q^*| \leq k$, is

$$\pi^{\leq}(n \mid k) = \pi_{\sigma_k}(n).$$

Proof: The proof is immediate, because we have that for $|n| = k$

$$\pi(n \mid k) = \frac{\mathbb{P}(Q^* = n, |n| = k)}{\mathbb{P}(|Q^*| = k)} = \frac{c_s \prod_{j=0}^{k} s_{j-1} \prod_{i \in V} \left(\frac{\lambda_i}{\mu_i}\right)^{n_i}}{\sum_{|n'|=k} c_s \prod_{j=0}^{k} s_{j-1} \prod_{i \in V} \left(\frac{\lambda_i}{\mu_i}\right)^{n'_i}} = \frac{\prod_{i \in V} \left(\frac{q_i}{\mu_i}\right)^{n_i}}{\sum_{|n'|=k} \prod_{i \in V} \left(\frac{q_i}{\mu_i}\right)^{n'_i}},$$

and, for $|n| \leq k$, provided that $s_i = 1$ for all $i = 0, 1, \ldots, k-1$, we have that

$$\pi^{\leq}(n \mid k) = \frac{\mathbb{P}(Q^* = n, |n| \leq k)}{\mathbb{P}(|Q^*| \leq k)} = \frac{c_s \prod_{j=0}^{|n|} s_{j-1} \prod_{i \in V} \left(\frac{\lambda_i}{\mu_i}\right)^{n_i}}{\sum_{|n'| \leq k} c_s \prod_{j=0}^{|n'|} s_{j-1} \prod_{i \in V} \left(\frac{\lambda_i}{\mu_i}\right)^{n'_i}} = \frac{\prod_{i \in V} \left(\frac{\lambda_i}{\mu_i}\right)^{n_i}}{\sum_{|n'| \leq k} \prod_{i \in V} \left(\frac{\lambda_i}{\mu_i}\right)^{n'_l}} = \pi_{\sigma_k}(n).$$

□

4.3 Equilibrium strategies

As already mentioned, we want to compute the expected profit function $P(k)$ (see Eq. (4.1)) of a joining customer that at her arrival observes k customers in the system. Thus, knowing the partial information, the tagged customer joins the system if $P(k) > 0$ and balks if $P(k) < 0$, while for $P(k) = 0$, the two actions are immaterial. We prove that such a strategy is an equilibrium threshold strategy.

We recall that, from the PASTA property (Theorem 1.12), a customer observes the state Q^* upon her arrival. We also recall that $T_i(k)$ is the expected sojourn time at queue $i \in V$, knowing the partial information $|Q^*| = k$.

We define $T_i^{\leq}(k) = \mathbb{E}_{\sigma_{k+1}}[S_i \mid |Q^*| \leq k]$ as the expected sojourn time at queue i of a joining customer under the strategy σ_{k+1}, where S_i is the random variable representing the sojourn time at queue i. The following lemma gives some properties about $T_i(k)$ and $T_i^{\leq}(k)$.

Lemma 4.3 (Properties of $T_i(k)$ and $T_i^{\leq}(k)$)**.** For any $i \in V$, the expected sojourn time $T_i(k)$ is independent of the strategy s adopted by all customers and it holds that

$$\mathbb{E}_{\sigma_{k+1}}[S_i \mid |Q^*| \leq k-1] = \mathbb{E}_{\sigma_k}[S_i \mid |Q^*| \leq k-1] =: T_i^{\leq}(k-1).$$

Proof: For any queue $i \in V$, we have that

$$T_i(k) = \sum_{|n|=k} \bar{T}_i(n+e_1)\pi(n \mid k),$$

where $\bar{T}_i(n+e_1)$ is the expected sojourn time at queue i of a joining customer observing the state $n \in \mathbb{Z}_+^{\nu}$ at her arrival and $\pi(n \mid k)$ is the conditional joint distribution, which does not depend on the strategy s, as proved in Corollary 4.2. So, also the expected sojourn time $T_i(k)$ does not depend on the strategy s.

For the second part of the statement, we have that, under a strategy s, with $s_i = 1$ for all $i = 0, 1, \ldots, k-1$,

$$\mathbb{E}_s[S_i \mid |Q^*| \leq k-1] = \sum_{j=0}^{k-1} T_i(j) \sum_{|n|=j} \pi^{\leq}(n \mid k-1),$$

since both $T_i(j)$ and $\pi^{\leq}(n \mid k)$ don't depend on s, as proved in Corollary 4.2. So

$$\mathbb{E}_{\sigma_{k+1}}[S_i \mid |Q^*| \leq k-1] = \mathbb{E}_{\sigma_k}[S_i \mid |Q^*| \leq k-1] =: T_i^{\leq}(k-1).$$

□

At this point, we compute $T_i(k)$ and then the expected profit function, which easily comes from Eq. (4.1). The proof of Theorem 4.4 use the Little's law to compute $T_i^{\leq}(k)$ and then, thanks to Lemma 4.3, the expression of $T_i(k)$ is recovered (see Remark 1.24).

Theorem 4.4 (Expected sojourn time)**.** For $i \in V$, the expected sojourn time at queue i is

$$T_i(k) = \frac{\sum\limits_{|n|=k+1} n_i \prod\limits_{\ell \in V} \left(\frac{q_\ell}{\mu_\ell}\right)^{n_\ell}}{\sum\limits_{|n|=k} \prod\limits_{\ell \in V} \left(\frac{q_\ell}{\mu_\ell}\right)^{n_\ell}}, \quad k = 0, 1, 2, \ldots$$

Proof: Let

$$x_j = \sum_{|n|=j} \prod_{\ell \in V} \left(\frac{\lambda_\ell}{\mu_\ell}\right)^{n_l},$$
$$y_j(i) = \sum_{|n|=j} n_i \prod_{\ell \in V} \left(\frac{\lambda_\ell}{\mu_\ell}\right)^{n_\ell}.$$

As defined in Remark 3.1, let σ_{k+1} denote the threshold strategy under which customers join the network if and only if the total number of customers in the network is less than $k+1$. Let $T_i^{\leq}(k)$ be the expected sojourn time at queue $i \in V$ incurred by a joining customer under the threshold strategy σ_{k+1}.

Under the threshold strategy σ_{k+1}, the joining probability of an arbitrary arriving customer is

$$\mathbb{P}(|Q^*| \leq k) = \sum_{|n|\leq k} \pi_{\sigma_{k+1}}(n) = \frac{\sum\limits_{j=0}^{k} \sum\limits_{|n|=j} \prod\limits_{\ell \in V} \left(\frac{\lambda_\ell}{\mu_\ell}\right)^{n_\ell}}{\sum\limits_{j=0}^{k+1} \sum\limits_{|n|=j} \prod\limits_{\ell \in V} \left(\frac{\lambda_\ell}{\mu_\ell}\right)^{n_\ell}} = \frac{\sum\limits_{j=0}^{k} x_j}{\sum\limits_{j=0}^{k+1} x_j},$$

where we have used Eq. (4.5) with $s = \sigma_{k+1}$. Thus, under the threshold strategy σ_{k+1}, the arrival rate of joining customers in the system, denoted by $\bar{\lambda}$, is

$$\bar{\lambda} = \lambda \frac{\sum_{j=0}^{k} x_j}{\sum_{j=0}^{k+1} x_j}.$$

For $i \in V$, the expectation of Q_i^*, under the threshold strategy σ_{k+1}, is

$$\mathbb{E}[Q_i^*] = \sum_{m=0}^{+\infty} m \sum_{\substack{n \in \mathbb{Z}_+^M \\ n_i = m}} \pi_{\sigma_{k+1}}(n) = \sum_{n \in \mathbb{Z}_+^M} n_i \, \pi_{\sigma_{k+1}}(n) = \frac{\sum\limits_{j=0}^{k+1} \sum\limits_{|n|=j} n_i \prod\limits_{\ell \in V} \left(\frac{\lambda_\ell}{\mu_\ell}\right)^{n_\ell}}{\sum\limits_{j=0}^{k+1} \sum\limits_{|n|=j} \prod\limits_{\ell \in V} \left(\frac{\lambda_\ell}{\mu_\ell}\right)^{n_\ell}} = \frac{\sum\limits_{j=0}^{k+1} y_j(i)}{\sum\limits_{j=0}^{k+1} x_j}.$$

From Little's law, as defined in (1.8), with $L = \mathbb{E}[Q_i^*]$, $T = T_i^{\leq}(k)$ and average arrival rate $\bar{\lambda}$, we have that

$$T_i^{\leq}(k) = \frac{\mathbb{E}\left[Q_i^*\right]}{\bar{\lambda}} = \frac{\sum_{j=0}^{k+1} y_j(i)}{\lambda \sum_{j=0}^{k} x_j}. \tag{4.7}$$

Let us note that we use the arrival rate $\bar{\lambda}$ for all node $i \in V$ because the tagged customer observes the system at her arrival, without knowing which path she will take.

Under the threshold strategy σ_{k+1}, $k \geq 1$, the probability that a joining customer has the information that there are less than k customers in the network is

$$\mathbb{P}(|Q^*| < k \mid |Q^*| \leq k) = \frac{\mathbb{P}(|Q^*| < k)}{\mathbb{P}(|Q^*| \leq k)} = \frac{\sum_{j=0}^{k-1} x_j}{\sum_{j=0}^{k+1} x_j} \cdot \frac{\sum_{j=0}^{k+1} x_j}{\sum_{j=0}^{k} x_j} = \frac{\sum_{j=0}^{k-1} x_j}{\sum_{j=0}^{k} x_j}$$

and the probability that a joining customer has the information that there are exactly k customers in the network is

$$\mathbb{P}(|Q^*| = k \mid |Q^*| \leq k) = \frac{\mathbb{P}(|Q^*| = k)}{\mathbb{P}(|Q^*| \leq k)} = \frac{x_k}{\sum_{j=0}^{k+1} x_j} \cdot \frac{\sum_{j=0}^{k+1} x_j}{\sum_{j=0}^{k} x_j} = \frac{x_k}{\sum_{j=0}^{k} x_j}.$$

Therefore, we have that, for $k = 1, 2, \ldots,$

$$T_i^{\leq}(k) = \frac{\sum_{j=0}^{k-1} x_j}{\sum_{j=0}^{k} x_j} \mathbb{E}_{\sigma_{k+1}}[S_i \mid |Q^*| \leq k-1] + \frac{x_k}{\sum_{j=0}^{k} x_j} \mathbb{E}_{\sigma_{k+1}}[S_i \mid |Q^*| = k],$$

which, from Lemma 4.3, turns out to be equal to

$$T_i^{\leq}(k) = \frac{\sum_{j=0}^{k-1} x_j}{\sum_{j=0}^{k} x_j} T_i^{\leq}(k-1) + \frac{x_k}{\sum_{j=0}^{k} x_j} T_i(k). \tag{4.8}$$

Substituting (4.7) into (4.8), we obtain that

$$T_i(k) = \frac{y_{k+1}(i)}{\lambda x_k} = \frac{\sum\limits_{|n|=k+1} n_i \prod\limits_{\ell \in V} \left(\frac{q_\ell}{\mu_\ell}\right)^{n_\ell}}{\sum\limits_{|n|=k} \prod\limits_{\ell \in V} \left(\frac{q_\ell}{\mu_\ell}\right)^{n_\ell}}, \quad k = 1, 2, \ldots$$

This also holds for $k = 0$, because we have that

$$T_i(0) = \frac{q_i}{\mu_i},$$

that is, knowing that no customers are in front of the tagged one, the expected sojourn time at queue i is given by the probability of arriving at queue i multiplied by the average service time at queue i. $\square$

We observe that, for all $i \in V$, $T_i(k)$ is strictly increasing and goes to $+\infty$ as $k \to +\infty$. It means that the expected sojourn time at queue $i \in V$ of a tagged customer grows, as the number of customers k grows, and it goes to $+\infty$, if the number of customers in the system goes to $+\infty$. This result is stated in the following theorem.

Theorem 4.5 (Monotonicity of the expected sojourn time)**.** For $i \in V$, the expected sojourn time at queue i, $T_i(k)$, is strictly increasing and

$$\lim_{k \to +\infty} T_i(k) = +\infty.$$

Proof: To prove the first part of the statement, we want to show that, for $i \in V$,

$$\frac{T_i(k)}{T_i(k-1)} > 1, \quad \forall k \in \mathbb{Z},\ k \geq 1.$$

We denote $a_\ell := q_\ell/\mu_\ell$ for $\ell \in V$. We observe that

$$\begin{aligned}\sum_{|n|=k+1} \prod_{\ell\in V} a_\ell^{n_\ell} &= \sum_{j\in V} a_j \sum_{|n|=k} \prod_{\ell\in V} a_\ell^{n_\ell}\\ \sum_{|n|=k+1} n_i \prod_{\ell\in V} a_\ell^{n_\ell} &= \sum_{j\in V} a_j \sum_{|n|=k} n_i \prod_{\ell\in V} a_\ell^{n_\ell} + a_i \sum_{|n|=k} \prod_{\ell\in V} a_\ell^{n_\ell}.\end{aligned} \tag{4.9}$$

Then, using the previous observation in (4.9) and Theorem 4.4, we have that

$$\begin{aligned}\frac{T_i(k)}{T_i(k-1)} &= \frac{\sum\limits_{|n|=k+1} n_i \prod\limits_{\ell\in V} a_\ell^{n_\ell}}{\sum\limits_{|n|=k} \prod\limits_{\ell\in V} a_\ell^{n_\ell}} \cdot \frac{\sum\limits_{|n|=k-1} \prod\limits_{\ell\in V} a_\ell^{n_\ell}}{\sum\limits_{|n|=k} n_i \prod\limits_{\ell\in V} a_\ell^{n_\ell}}\\ &= \frac{\sum\limits_{j\in V} a_j \sum\limits_{|n|=k} n_i \prod\limits_{\ell\in V} a_\ell^{n_\ell} + a_i \sum\limits_{|n|=k} \prod\limits_{\ell\in V} a_\ell^{n_\ell}}{\sum\limits_{j\in V} a_j \sum\limits_{|n|=k-1} \prod\limits_{\ell\in V} a_\ell^{n_\ell}} \cdot \frac{\sum\limits_{|n|=k-1} \prod\limits_{\ell\in V} a_\ell^{n_\ell}}{\sum\limits_{|n|=k} n_i \prod\limits_{\ell\in V} a_\ell^{n_\ell}}\\ &= 1 + \frac{a_i \sum\limits_{|n|=k} \prod\limits_{\ell\in V} a_\ell^{n_\ell}}{\sum\limits_{j\in V} a_j \sum\limits_{|n|=k} n_i \prod\limits_{\ell\in V} a_\ell^{n_\ell}} > 1.\end{aligned}$$

Regarding the second part of the statement, we denote $c_i := a_i/\sum\limits_{j\in V} a_j$. Then, we have that $T_i(k)/T_i(k-1)$ can be bounded by $1 + c_i k^{-1}$ from below, since

$$\frac{T_i(k)}{T_i(k-1)} = 1 + \frac{a_i \sum\limits_{|n|=k} \prod\limits_{\ell\in V} a_\ell^{n_\ell}}{\sum\limits_{j\in V} a_j \sum\limits_{|n|=k} n_i \prod\limits_{\ell\in V} a_\ell^{n_\ell}} \geq 1 + \frac{a_i \sum\limits_{|n|=k} \prod\limits_{\ell\in V} a_\ell^{n_\ell}}{k \sum\limits_{j\in V} a_j \sum\limits_{|n|=k} \prod\limits_{\ell\in V} a_\ell^{n_\ell}} = 1 + c_i \,\frac{1}{k}, \tag{4.10}$$

where we use the fact that $n_i \leq k$ for any $n \in \mathbb{Z}_+^\nu$ such that $|n| = k$.

Hence, using Eq. (4.10), the limit of $T_i(k)/T_i(0)$ becomes

$$\lim_{k\to+\infty} \frac{T_i(k)}{T_i(0)} = \lim_{k\to+\infty} \prod_{j=1}^{k} \frac{T_i(j)}{T_i(j-1)} \geq \lim_{k\to+\infty} \prod_{j=1}^{k} \left(1 + c_i \frac{1}{j}\right) > 1 + c_i \lim_{k\to+\infty} \sum_{j=1}^{k} \frac{1}{j} = +\infty,$$

and this concludes the proof, since $T_i(0)$ is constant and equal to q_i/μ_i. □

From Theorem 4.4 and using Eq. (4.1), we can build the expected profit function $P(k)$, which turns out to be equal to

$$P(k) = R - \sum_{i\in V} C_i T_i(k) = R - \sum_{i\in V} C_i \frac{\sum\limits_{|n|=k+1} n_i \prod\limits_{\ell\in V} \left(\frac{q_\ell}{\mu_\ell}\right)^{n_\ell}}{\sum\limits_{|n|=k} \prod\limits_{\ell\in V} \left(\frac{q_\ell}{\mu_\ell}\right)^{n_\ell}}, \quad k = 0, 1, 2, \ldots$$

We observe that $P(k)$ is independent of the strategy s and the arrival rate λ. Furthermore, from Theorem 4.5, we have that $P(k)$ is strictly decreasing in k and $\lim_{k\to+\infty} P(k) = -\infty$, which means that the more the customers in the system at the arrival of a customer, the lower her expected profit function.

Thanks to this observation, we have that the lower extreme $\inf\{k \in \mathbb{Z}_+ : P(k) < 0\}$ exists and is finite. In the following theorem, we show which is the strategy that all customers have to use to get an optimal payoff. In particular, we have that the equilibrium strategies are threshold strategies.

Theorem 4.6 (Equilibrium strategies)**.** Let K be the threshold given by

$$K = \inf\{k \in \mathbb{Z}_+ : P(k) < 0\}.$$

If $P(K-1) > 0$, then the threshold strategy σ_K is the only equilibrium strategy. If $P(K-1) = 0$, then all the threshold strategies $(1-p)\sigma_{K-1} + p\sigma_K$, with $p \in [0,1]$, are the only equilibrium strategies.

Proof: Let $F_{s^1,s^2}(k)$ be the payoff for a customer that chooses to follow the strategy s^1, when everyone else chooses to follow the strategy s^2. In this case, the payoff function is equal to $P(k)$ if the strategy s^1 says to enter the system, 0 otherwise, that is

$$F_{s^1,s^2}(k) = P(k)X(k),$$

where $X(k)$ is a Bernoulli random variable of parameter s^1_k.

We have that a strategy $\bar{s}$ is an equilibrium strategy if and only if $F_{\bar{s},s^2}(k) \geq F_{s^1,s^2}(k)$, for all $k \in \mathbb{N}$ and for all strategies s^1, s^2, and, by the definition of the payoff function, it is equivalent to

$$\begin{aligned} \bar{s}_k &= 1 \quad \text{for } k \text{ with } P_s(k) > 0 \quad \text{and} \\ \bar{s}_k &= 0 \quad \text{for } k \text{ with } P_s(k) < 0. \end{aligned} \tag{4.11}$$

So, the threshold strategies in the statement are the only strategies that satisfy equations (4.11), and so they are the only equilibrium strategies for the tree network. □

Now, we consider queue $j \in V \setminus \{1\}$ and analyze what happens to the system if the service time at queue j goes to infinity. We compute the limit of $T_i(k)$ as μ_j goes to $+\infty$ for any $i \in V$. Doing the limit means supposing that, as soon as a customer reaches queue j, she immediately completes the service and moves to one of the next queues in $S^-(j)$ if $j \in V_1$, or leaves the system. So, we expect that the model behaves as a tree network with one less vertex and with edges connecting directly the vertex $S^+(j)$ to any vertex in $S^-(j)$. To prove this result, we compute the limit of $T_i(k)$ for any $i \in V$. We obtain that for $k = 0, 1, 2, \ldots,$

$$\begin{aligned} &\lim_{\mu_j\to\infty} T_j(k) = 0, \\ &\lim_{\mu_j\to\infty} T_i(k) = \frac{\sum\limits_{|n|=k+1} n_i \prod\limits_{\ell\in V\setminus\{j\}} \left(\frac{q_\ell}{\mu_\ell}\right)^{n_\ell}}{\sum\limits_{|n|=k} \prod\limits_{\ell\in V\setminus\{j\}} \left(\frac{q_\ell}{\mu_\ell}\right)^{n_\ell}}, \quad \text{for } i \neq j, \end{aligned} \tag{4.12}$$

where the sum over $|n| = k+1$ and $|n| = k$ is made for $n \in \mathbb{Z}_+^{\nu-1}$, obtained by removing the j-th entry.

To conclude this section, we consider an example of tree network, where the expected profit function is greatly simplified. We assume that all service rates and all costs for sojourn time are equal to each other, that is $\mu_i = \mu_j =: \mu$ and $C_i = C_j =: C$ for all $i, j \in V$. Then the expected profit function $P(k)$ becomes

$$P(k) = R - \frac{(k+1)C}{\mu} \cdot \frac{\sum\limits_{|n|=k+1} \prod\limits_{\ell \in V} q_\ell^{n_\ell}}{\sum\limits_{|n|=k} \prod\limits_{\ell \in V} q_\ell^{n_\ell}}. \tag{4.13}$$

.

4.4 Numerical computations

The easiest example of a tree network that is not a tandem network has an underlying out-tree $G(V,E)$ with $V = \{1,2,3\}$ and $E = \{(1,2),(1,3)\}$. First, we compare this tree network with a two-node tandem network. Then, we look at the behaviour of the equilibrium threshold for different values of costs, as the routing probabilities vary. We refer to the tree network with *model A* and to the two-node tandem network with *model B*.

To compare the two models, for *model A* we choose the following parameters: the routing probabilities are $p_{12} = p_{13} = 1/2$, which implies that $q_1 = 1$ and $q_2 = q_3 = 1/2$, and the service rates are all equal to μ. For the two-node tandem network (see Chapter 2), we choose $\mu_1 = \mu$, $\mu_2 = 2\mu$.

So, we are comparing the two queueing networks represented in Figure 4.2, where at queue 1 the service rates are equal to μ and, after completing service at queue 1, customers are routed, on the one hand, to queue 2 or queue 3 with the same probability of $1/2$, both with service rate μ, and, on the other hand, to queue 2 with a double service rate 2μ. In both systems, customers pay a cost C per unit of sojourn time at each queue.

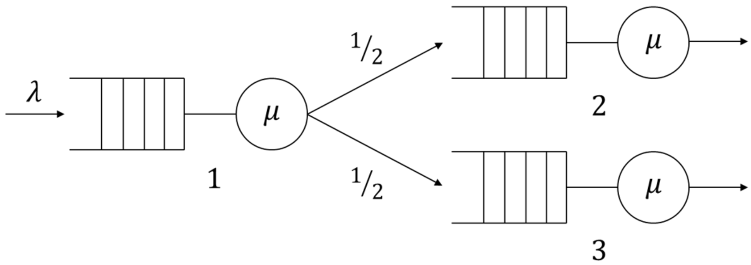
We choose to compare these two models, because we expect the equilibrium strategies to be similar to each other, due to the fact that, after queue 1, in *model A* customers are split into two different queues, while in *model B* customers are routed to the same queue, but with a double service rate.

For *model A* (see Chapter 2), the expected profit function is

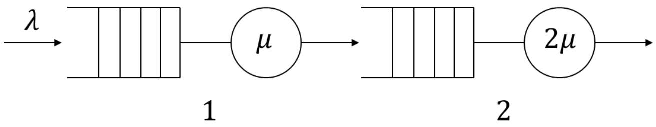
$$P_A(k) = R - \frac{(k+1)C}{\mu} \cdot \frac{2(k+1) + 2^{-(k+1)}}{2k + 2^{-k}},$$

while, for *model B* (see Eq. (4.13)), it is

$$P_B(k) = R - \frac{(k+1)C}{\mu} \cdot \frac{2^{k+2} - 1}{2^{k+2} - 2}.$$



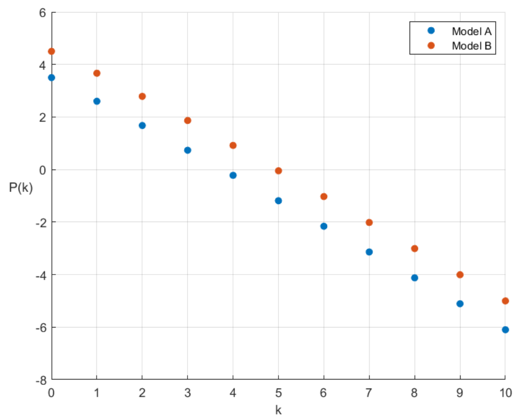
(a) *Model A*.



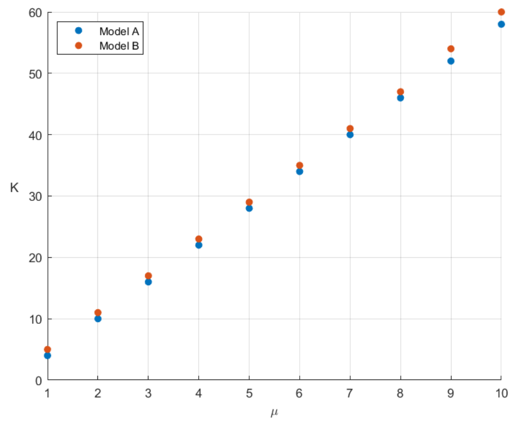
(b) *Model B*.

Figure 4.2: Comparison of *model A* with *model B*.

The functions $P_A(k)$ and $P_B(k)$ are graphically represented in Figure 4.3 (a), where we fix some parameters in order to see their behaviour. We observe that $P_A(k)$ is slightly lower than $P_B(k)$. In Figure 4.3 (b), however, we compare the equilibrium thresholds for the models, noting again that the thresholds for *model A* are lower than those for *model B*.



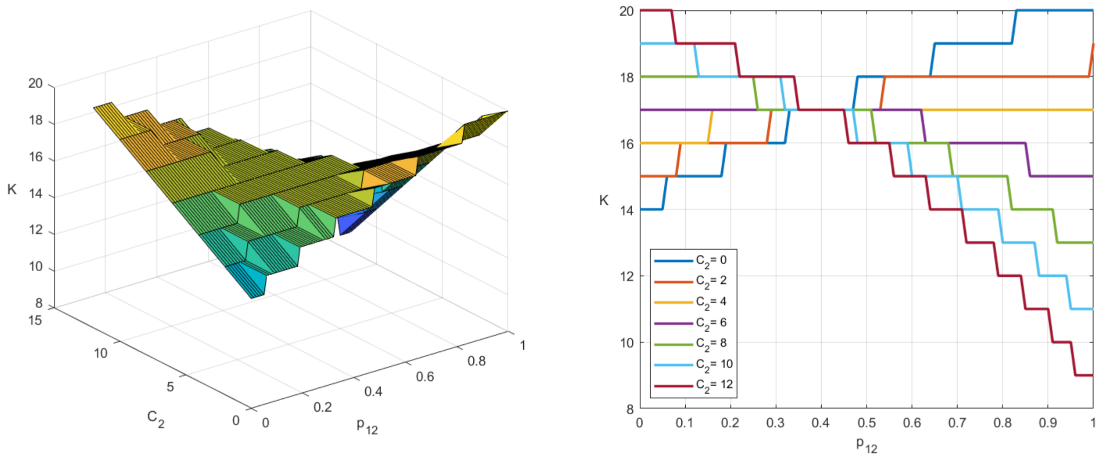
(a) Plot of $P_A(k)$ and $P_B(k)$ as k varies.



(b) Equilibrium thresholds as μ grows.

Figure 4.3: In these graphs, we have chosen the following parameters $R = 6, C = 1$ and, in figure (a), also $\mu = 1$.

Now, we take a closer look at *model A*, by varying the routing probabilities and the costs for sojourn time, in order to see how the equilibrium threshold changes. We fix the following parameters: service rates are $\mu_i = i$ for $i = 1, 2, 3$ and the reward is $R = 20$. We also fix $C_1 = 1$, $C_1 + C_2 + C_3 = 13$ and $p_{12} + p_{13} = 1$. On the contrary the parameters that vary are $C_2 \in [0, 12]$ (with $C_3 = 12 - C_2$) and $p_{12} \in [0, 1]$ (with $p_{13} = 1 - p_{12}$). Within this framework, the values of the equilibrium threshold are represented in Figure 4.4. We can observe that the thresholds are higher when the cost C_2 is low and the probability p_{12} is high or when the cost C_2 is high and the probability p_{12} is low, while thresholds are lower when C_2 and p_{12} are both high or low. A sort of saddle point is formed in the middle. In Appendix A, we include, as an example, the MATLAB code for the expected profit function of a customer who joins a tree network. For a complete collection of MATLAB codes, see [19].



(a) 3D-plot. (b) 2D-plot as p_2 varies.

Figure 4.4: Plot (a) and plot (b) show the values of the equilibrium threshold as the routing probability p_{12} and the cost C_2 vary, where $p_{13} = 1 - p_{12}$ and $C_1 = 1$, $C_3 = 12 - C_2$.

Chapter 5

Overtaking free condition

Abstract

In this chapter we show that, under the FIFO discipline and with exponential service times, a queueing network is overtaking free if and only if it is a tree network. Then, we show that there exist systems that are not overtaking free under the FIFO discipline, but they become so if the service discipline is changed. After that, we study an example of non overtaking free network under the FIFO discipline. Finally, we consider the $M/E_r/1$ model, where the distribution of service times, instead of being exponential, is Erlang.

5.1 Overtaking free networks

So far we have only considered queueing networks, where the place in each queue is determined by the order of customers' arrivals. This means that whoever arrives at a certain queue first, gets served first. Such service discipline is called *FIFO*, which stands for *first-in-first-out*. In this chapter, precisely in Section 5.2, we change the customer service discipline. We suppose that customers arriving at the system are labeled with increasing numbers, which means that those who arrive first at the system have lower numbers than those who arrive later. When the tagged customer arrives at a certain queue along the system, she is directly ahead of everyone that is tagged with a higher number than hers, even if someone of these arrived first. Within this service discipline, we have that the possible customers ahead of the tagged one are only those arrived at the system before her and, so, it doesn't need to take into account the presence of customers arrived at the system after her for the computation of the sojourn time. Such a service discipline is called *preemptive-resume label order discipline*.

We explain better the preemptive-resume label order discipline in the following example, graphically represented in Figure 5.1. We suppose that a Poisson process settles the arrivals at the system by sending them into two queues, called queue 1 and queue 2, and then they are all routed to the same queue, called queue 3. Let us suppose that the tagged customer is associated to number N and she is routed to queue 1. The next arriving customer will be labeled with $N + 1$ and so on. When customer N arrives to queue 3, she overtakes immediately all customers with greater labels, including, in case, also the one receiving the service. On the contrary, while customer N waits in queue 3, any customer, who is labeled with lower number and arrives from queue 2, overtakes her and receives the service before her. Thanks to this service discipline, customers with greater labels than that of the tagged customer don't affect the sojourn time of the tagged customer, while those with lower labels do, since they arrived at the system before her.

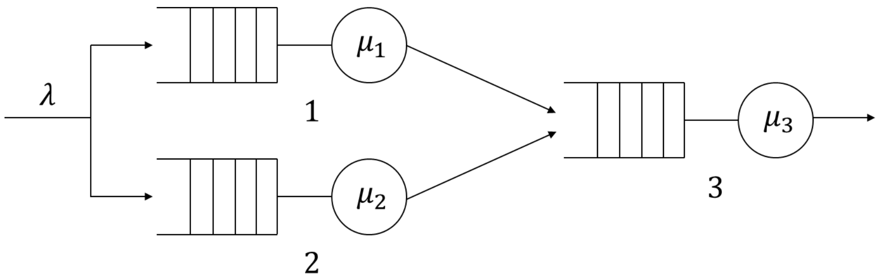


Figure 5.1: Example of queueing network, where, under the FIFO discipline, the network is not overtaking free, while, under the preemptive-resume label order discipline, the network is overtaking free.

In this section, we prove that, under a FIFO discipline and with exponential service times, a queueing network is overtaking free if and only if it is a tree network. We first give a formal definition of overtaking free network [20, 11] and tree network [6, 2], and then prove the statement.

We suppose that all joining customers arrive from a Poisson process and enter the system at queue 1. That is the transition probability from outside (node 0) to first node (node 1) is $p_{01} = 1$ and node 1 serves as the root of the tree. We also assume that there exists at least one node, from which customers leave the system, and at least one path from 1 to any node of the network, since otherwise such a node would be out of the network. We say that the underlying graph of a queueing network is the directed graph $G(V, E)$, where V is the set of nodes of the network and E is the set of edges (i, j) such that $p_{ij} > 0$.

Definition 5.1 (Overtaking free network)**.** A queueing network is overtaking free if and only if the sojourn time of a tagged customer at any queue doesn't depend on the distribution of the customers arrived at the system after her.

Definition 5.2 (Tree network)**.** A queueing network is a tree network if and only if its underlying graph $G(V, E)$ is an out-tree.

Theorem 5.3 (Characterization of overtaking free networks)**.** Under the FIFO discipline and with exponential service times at each queue, a queueing network is overtaking free if and only if it is a tree network.

Proof: First, we observe that according to an exponential distribution, there is a positive probability that a customer takes a time t to complete the service for any $t > 0$. We prove one implication at a time.

($\Leftarrow$) Let $G(V, E)$ be the underlying graph of a tree network, namely $G(V, E)$ is an out-tree. Then, from Lemma 1.29, we have that there is a unique path that connects the root node 1 to any vertex of the graph. So, since the service discipline is the FIFO one, the queuing network is overtaking free.

($\Rightarrow$) Let $G(V, E)$ be the underlying graph of an overtaking free network. Using Definition 1.28, we prove that $|E| = |V| - 1$, because by assumption we already have that for any node $v \in V$ there exists a directed path from node 1 to node v. Let us suppose by contradiction that there exists two different edges entering a vertex $v \in V \setminus \{1\}$. It means that there are two different paths connecting vertex 1 to vertex v, say w_1 and w_2. Then, since the service times are exponential and the discipline is FIFO, there is a positive probability such that a customer running along path w_2 can reach queue at node v and be in front of a customer, who arrived at the system before her and ran along path w_1. This is a contradiction with the definition of overtaking free condition. So, since at least one edge enters the vertex v, we have that any vertex has one only entering edge. As for vertex 1, if there exists an edge entering in 1, then there is also a cycle, which is a contradiction because in this way there would exist two different path connecting vertex 1 to any vertex. So, the edges are as many as the vertices minus 1, that is $|E| = |V| - 1$. □

We observe that the proof holds thanks to the assumption on the service discipline and the distribution of the service times, but if we consider other distributions of service times, provided that their support is $[0, +\infty)$, as, for example, the Erlang distribution, the theorem still holds true. However, in general, if we change the service discipline or the distribution of the service times, Theorem 5.3 may no longer be valid.

So, in conclusion, we have that, under a FIFO discipline and with exponential service times, in an overtaking free network, node 1 serves as the root of the tree, at which customers arrive, and all other nodes must have only one entering edge. So, it means that customers, after the completion of the service, can be divided into several queues, but each queue can accept customers from only one queue.

An example of non overtaking free queueing network with exponential service times, under a FIFO discipline, is represented in Figure 5.2. It is easy to see that the overtaking free condition is not satisfied since a customer, while waiting at queue 1, can be overtaken by a customer, who is directly routed to queue 2. As analyzed in [20], we have that in this example the sojourn times of a customer at queue 1 and at queue 2 are not independent. On the contrary, in an overtaking free network, we have that the sojourn times of a customer in different queues are independent. For the proof about independence when queues are in tandem, see [16, 17].

5.2 Preemptive-resume label order discipline

In this section we analyse a two-node tandem network, where customers can enter the system also at the second queue, as represented in Figure 5.2. This is an example of queueing network which is not overtaking free under the FIFO discipline, but it becomes so if the discipline is changed. Specifically, we will use the preemptive-resume label order discipline.

We consider a tandem network with two single server nodes with infinite buffers. We index the nodes by l, $l = 1, 2$. Service times are independent and exponentially distributed with mean μ_l^{-1}, that is the service rate at node l is μ_l.

We suppose that customers arrive to the system according to a Poisson process with rate λ and they are routed with probability $p_1 > 0$ to queue 1 and with probability p_2 to queue 2, where $p_1 + p_2 = 1$. That is, from Theorem 1.11, we can suppose that customers arrive to the system according to two independent Poisson processes of intensities $\lambda_1 := p_1\lambda$ and $\lambda_2 := p_2\lambda$. If customers arrive to the system from the first Poisson process, then they are routed to queue 1, otherwise they are routed directly to queue 2.

In this queueing network customers are served according to the order of their labels. That is when a customer arrive at the system, she is associated to a number and the service discipline is the preemptive-resume label order one, as defined previously in Section 5.1. The model is graphically represented in the following queueing network (Figure 5.2).

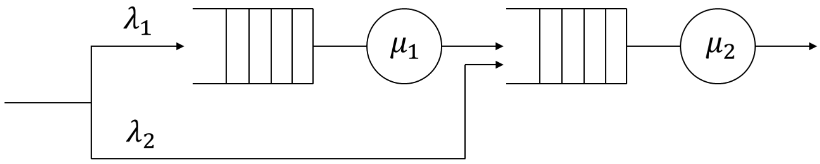


Figure 5.2: Two-node tandem network where customers arrive at the system following one of the two independent Poisson processes with intensities λ_1 and λ_2. In one case they are routed to queue 1, in the other one to queue 2. Service rates are μ_1 and μ_2 respectively.

We are interested in finding the strategy to be adopted by all arriving customers such that it is an equilibrium for customers arriving from the first Poisson process. As usual, the partial information is the number of customers in the system at the arrival of the tagged customer, who, therefore, doesn't know the exact number of customers in the first queue and in the second one, but only the total number.

From now on, we say arriving customer to mean a customer that arrive at the system from the Poisson process of intensity λ_1. So, the arriving customer, after receiving the partial information about the state of the system, decides whether to join or balk the system according to the strategy.

For our goal, we compute the expected profit $P(k)$ of an arriving customer that observes k customers in the system and joins. By maximizing the expected profit function, we can find the equilibrium threshold. The expected profit function has the following form

$$P(k) = R - C_1 T_1(k) - C_2 T_2(k), \tag{5.1}$$

where $R \geq C_1/\mu_1 + C_2/\mu_2$ is the reward for joining the system, C_l is the cost for unit of sojourn time at queue l and $T_l(k)$ is the expected sojourn time at queue l, knowing the partial information.

We assume that all arriving customers use the same strategy. By definition 3.12, an equilibrium strategy for this model requires the following actions: to join if $P(k) > 0$, to balk if $P(k) < 0$. We will see that such a strategy is a threshold strategy.

As for Chapter 2, the random vector (Q_1, Q_2) denotes the state of the system: $(Q_1, Q_2) = (n, m)$ means that n customers are in the first queue and m customers in the second queue. Let $T_l(n, m)$ be the expected sojourn time spent at queue l by an arriving customer that joins a system being in state $(n-1, m)$, and $T(n, m) = T_1(n, m) + T_2(n, m)$ the total expected sojourn time. Then, we have that

$$T_1(n, m) = \frac{n}{\mu_1}, \quad n, m \geq 0,$$

and, applying a first step analysis, we get that

$$T(n, m) = \frac{1}{\mu_1 + \mu_2} + \frac{\mu_1}{\mu_1 + \mu_2} T(n-1, m+1) + \frac{\mu_2}{\mu_1 + \mu_2} T(n, m-1), \quad n, m > 0, \tag{5.2}$$

which implies that

$$T_2(n, m) = \frac{\mu_1}{\mu_1 + \mu_2} T_2(n-1, m+1) + \frac{\mu_2}{\mu_1 + \mu_2} T_2(n, m-1), \quad n, m > 0.$$

The boundary conditions are

$$T(0, m) = T_2(0, m) = \frac{m}{\mu_2}, \quad m \geq 0,$$
$$T(n+1, 0) = \frac{1}{\mu_1} + T(n, 1) \quad \text{and} \quad T_2(n+1, 0) = T_2(n, 1), \quad n \geq 0.$$

We observe that the previous equations are the same ones that hold also for the two-node tandem network described in Chapter 2. In particular we observe that Eq. (5.2) holds because we don't consider the possibility of going from state (n, m) to state $(n+1, m)$ or state $(n, m+1)$ for the computation of the expected sojourn time, since they don't affect the expected sojourn time of a customer in position (n, m). Indeed the uniform Markov chain that described the position of a customer, from when she joins until she leaves the system, is the same as the one used in the two-node tandem network without entries (see Figure 2.3).

On the contrary, the stationary distribution of the number of customers in the system is different. Let (Q_1^*, Q_2^*) be the stationary state of the system and $Q^* = Q_1^* + Q_2^*$ the total stationary number of customers in the system. Let $\pi_K(n,m)$ be the stationary distribution of the state of the system, conditioned on $Q^* \leq K$. It means that the model is assumed to be a semiopen Jackson network with an overall buffer capacity of K (see Section 1.1). From the traffic equation (1.1), we obtain that the arrival rate at queue 1 and at queue 2 are $p_1\lambda$ and λ respectively. So, the stationary distribution, given by Theorem 1.3, is equal to

$$\pi_K(n,m) = c_K \left(\frac{p_1\lambda}{\mu_1}\right)^n \left(\frac{\lambda}{\mu_2}\right)^m, \quad n+m \leq K, \tag{5.3}$$

where $c_K^{-1} = \sum_{n+m\leq K} \left(\frac{p_1\lambda}{\mu_1}\right)^n \left(\frac{\lambda}{\mu_2}\right)^m$ is the normalization constant.

We denote with $\pi(n, k-n) = \mathbb{P}(Q_1^* = n | Q^* = k)$ the stationary distribution conditioned on the partial information, where we omit the subindex K because it doesn't depend on the threshold K. Then, from formula (5.3), we have that

$$\pi(n,k-n) = \begin{cases} \mu_1^{k-n}(p_1\mu_2)^n(\mu_1 - p_1\mu_2)/(\mu_1^{k+1} - (p_1\mu_2)^{k+1}), & \mu_1 \neq p_1\mu_2 \\ 1/(1+k), & \mu_1 = p_1\mu_2 \end{cases}. \tag{5.4}$$

Thanks to the conditional stationary distribution (5.4), we can compute the expected sojourn time at queue l, $T_l(k)$, for a customer that enters the system containing k customers. We obtain that, for $\mu_1 \neq p_1\mu_2$,

$$\begin{aligned} T_1(k) &= \frac{1}{\mu_1 - p_1\mu_2} - \frac{k+1}{\mu_1}\frac{(p_1\mu_2)^{k+1}}{\mu_1^{k+1} - (p_1\mu_2)^{k+1}}, \\ T_2(k) &= \frac{1}{p_1\mu_2 - \mu_1} - \frac{k+1}{p_1\mu_2}\frac{\mu_1^{k+1}}{(p_1\mu_2)^{k+1} - \mu_1^{k+1}}, \end{aligned}$$

and, for $\mu_1 = p_1\mu_2$,

$$T_l(k) = \frac{1}{\mu_1}\left(1 + \frac{k}{2}\right), \quad l = 1, 2.$$

We note that the formulas of $T_1(k)$ and $T_2(k)$ are symmetric if we exchange the terms μ_1 and $p_1\mu_2$. These expressions allow us to write a closed formula also for the expected profit function $P(k)$, as defined in (5.1).

At this point, since the tagged arriving customer wants to maximize her payoff function given by the expected profit function, if she enters, and 0, if she balks, the strategy to adopt says to join if $P(k)$ is strictly positive and to balk if $P(k)$ is strictly negative. Such a strategy is indeed an equilibrium strategy, consistent with Definition 1.17. Therefore, since the expected profit function $P(k)$ is strictly decreasing and goes to $-\infty$ as $k \to +\infty$, the equilibrium strategy is the K-threshold strategy, where the threshold K is given by

$$K = \arg\min\{k \in \mathbb{Z}_+ : P(k) < 0\}.$$

So, if all customers use the K-threshold strategy, it means that an arriving customer observes the number of users in the system and if this number is less than K, she joins, otherwise she balks.

5.3 Example of non overtaking free network

In this section we show an example of non overtaking free queueing network in order to see the different approach in the analysis of this type of model. For overtaking free queueing networks the sojourn time of the tagged customer is affected only by those ones that join the system before her arrival. But, this property doesn't hold when overtaking is admitted. It means that we need to take into account also the presence of customers arriving after the tagged one. This is the reason why studying non overtaking queueing network turns out to be more complicated.

The example of non overtaking free system is the following: we consider a queueing network with a single node, like in the $M/M/1$ model (see Section 1.5), but we assume that some arriving customers, the so-called *ordinary customers*, queue up behind everyone, while other arriving customers, the so-called *priority customers*, can overtake customers in line and queue up in front of the others, preempting the customer, who might be in service. Ordinary customers follow a FIFO discipline, while priority customers follow a LIFO discipline with preemption. We assume that that service, when resumed, is continued from the point it was interrupted.

So, we suppose that each customer arrives at the system according to a Poisson process with rate λ and can be an ordinary customer with probability $p_1 > 0$ and a priority customer with probability p_2, where $p_1 + p_2 = 1$. That is, from Theorem 1.11, we can suppose that customers arrive to the system according to two independent Poisson processes of intensity $\lambda_1 := p_1\lambda$ and $\lambda_2 := p_2\lambda$. If customers arrive to the system from the first Poisson process, they are ordinary customers, otherwise they are priority customers. We also suppose that the service times are independent and exponentially distributed with rate μ. The model is graphically represented in Figure 5.3.

We are interested to find the equilibrium strategy to be adopted by all arriving ordinary customers, assuming that priority customers always join. When an ordinary customer arrives at the system, she receives the partial information about the total number of customer in the system and decides whether to join or balk according to her strategy.

As usual, let $P(k)$ be the expected profit of a joining ordinary customer after observing k customers in the system. Then

$$P(k) = R - CT(k), \tag{5.5}$$

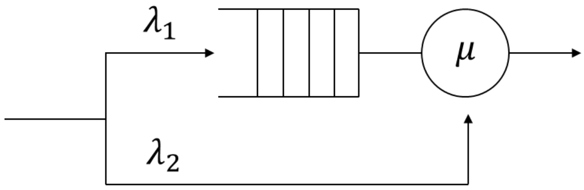


Figure 5.3: A single queue where customers arrive to the system following one of the two independent Poisson processes. In one case they are ordinary customers, in the other one priority customers. Service rate is μ.

where R is the reward for joining the system, C is the cost for unit of sojourn time and $T(k)$ is the expected sojourn time of a joining ordinary customer under the partial information.

Let us compute the expected sojourn time $T(k)$. For $k = 0$, we have that the arriving ordinary customer observes that the queue is empty and, if she joins, she immediately begins the service. But, if a priority customer arrives when the tagged customer is still in service, she preempts that customer and starts to be served. Only when there are no more priority customers in the queue, the tagged customer can resume her service from the point it was interrupted. The same happens for any priority customer arriving at the queue before the ordinary customer ends her service.

Actually, $T(0)$ is also the expected sojourn time of a priority customer that joins the system, because, by overtaking all customers in the queue, she immediately accesses the service and she is preempted only by other priority customers. As a consequence of this, any joining priority customer, knowing or not the partial information, has a profit function equal to $R - CT(0)$. So we suppose that $R > CT(0)$ to ensure that any priority customer and an ordinary customer, that observes an empty system, join.

We assume that $\mu > \lambda_2$, then for $k = 0$, we have that

$$T(0) = \frac{1}{\mu - \lambda_2}. \tag{5.6}$$

Indeed, $T(0)$ coincides with the expected sojourn time of the unobservable $M/M/1$ model (see Section 1.5) because, using the Little's law, the expected sojourn time is computed dividing the average number of customers by the arrival rate, each of which doesn't depend on the service discipline.

To compute $T(k)$ for $k \geq 1$, we use the following recursive equation, obtained with a first step analysis strategy

$$T(k) = \frac{1}{\mu + \lambda_2} + \frac{\lambda_2}{\mu + \lambda_2} T(k+1) + \frac{\mu}{\mu + \lambda_2} T(k-1), \quad \text{for all} \quad k = 1, 2, \dots \tag{5.7}$$

The solution of the recursive equation (5.7) with the boundary condition (5.6) is

$$T(k) = \frac{k+1}{\mu - \lambda_2}, \quad \text{for all} \quad k = 0, 1, 2, \dots, \tag{5.8}$$

and, so, the expected profit function becomes

$$P(k) = R - \frac{C(k+1)}{\mu - \lambda_2},$$

Hence, the equilibrium strategy is a threshold strategy with threshold K given by

$$K = \inf\{k \in \mathbb{Z}_+ : P(k) < 0\} = \left\lfloor \frac{R(\mu - \lambda_2)}{C} \right\rfloor. \tag{5.9}$$

If $P(K-1) > 0$, the pure K-threshold strategy is the only equilibrium strategy for this model, while, if $P(K-1) = 0$, any threshold strategy with threshold $x = K - p$ with $p \in [0, 1]$ is an equilibrium strategy and, as a consequence of this, there are both pure and mixed threshold strategies.

We observe that the equilibrium threshold in the Naor's model, computed in (1.6), is greater than the equilibrium threshold (5.9). This observation is consistent with the fact that the expected sojourn time of an ordinary customer in this model is affected by the priority overtaking customers and, so, it is larger than the same in the Naor's model.

5.4 $M/E_r/1$ model

In this section, we consider a model with one only queue, at which customers arrive according to a Poisson process of intensity λ and are served according to a FIFO discipline. But, differently from all previous systems, in this case the service times are independent and Erlang distributed. Let g denotes the service time, then $g \sim \text{Erlang}(r, \mu)$, with $r \in \mathbb{N}$. Assuming an Erlang distribution for the service times allows for the interpretation of one service time as a sequence of r independent phases, each exponentially distributed with parameter μ. Until a given customer in service leaves the queue, no other customer can start her service. This system is denoted by $M/E_r/1$. For the notation, we refer to [1, 10].

A state of this system can be represented by the number of customers in the system and the remaining number of phases of the customer in service. The state (n, m) means that in the system there are n customers and the one in service is missing m phases before leaving the system. The flow diagram in Figure 5.4 (a) describes this system.

An easier way to represent a state of the system is by means of the total number of remaining phases of work in the system. Indeed, the state (n, m) correspond to $(n-1) \cdot r + m$ phases of work to be done by the server. On the other hand, $N > 0$ phases of work in the system correspond to the state

$$\left(\left\lfloor \frac{N-1}{r} \right\rfloor + 1, N - \left\lfloor \frac{N-1}{r} \right\rfloor \cdot r \right).$$

The Markov chain associated to the number of phases of work in the system has state space $\{0, 1, 2, \dots\}$, where state N corresponds to N phases of work in the system, and it is represented in Figure 5.4 (b). The two ways to represent the states are equivalent, but the second one is easier for computations. So, we will use the second one.

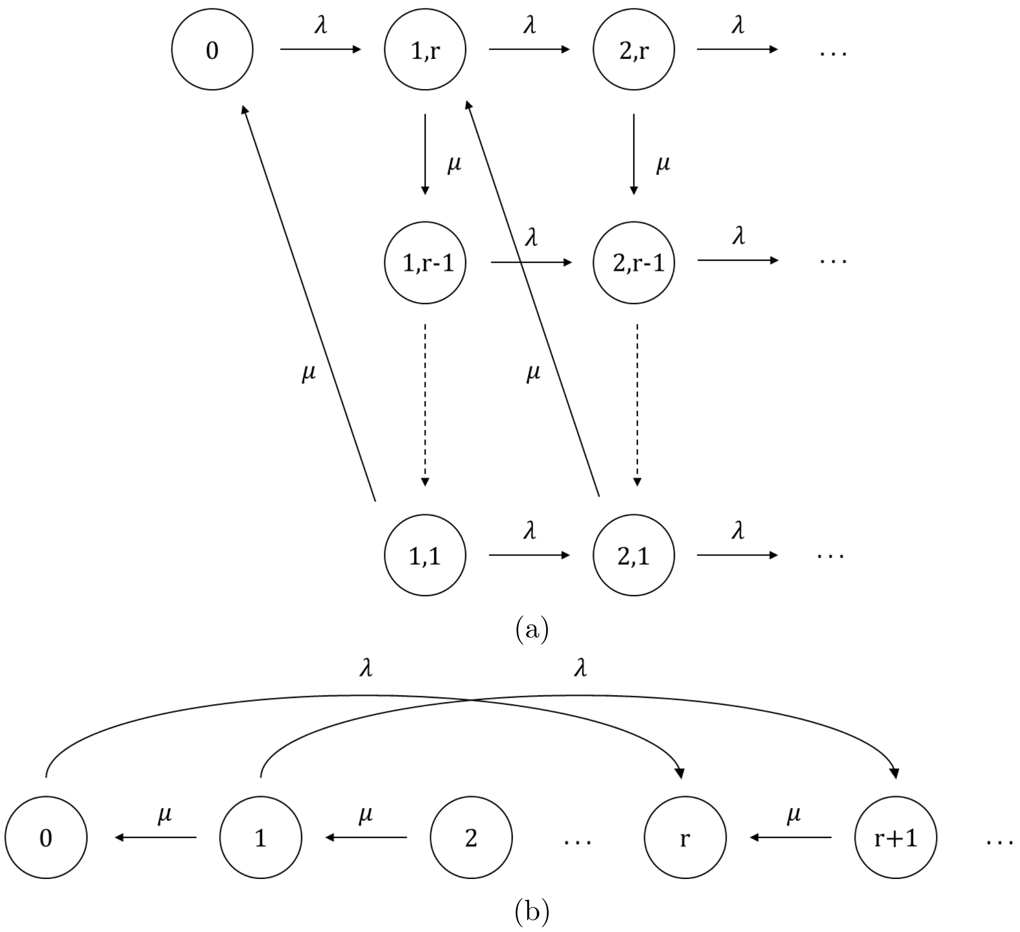


Figure 5.4: Transition rate diagrams of the $M/E_r/1$ model with arrival rate λ and Erlang service times with parameters r and μ.

We note that this system is equal to the $M/M/1$ model, except for the distribution of the service times, which are Erlang instead of exponential. We want to find an equilibrium strategy, if there is one, such that, all customers, using that strategy, get the maximum expected profit. Customers are informed about the total number of users in the system, say k, and the expected profit function $P(k)$ depends on this information, being defined by

$$P(k) = R - CT(k),$$

where R is the reward, C the unitary cost for sojourn time and $T(k)$ the expected sojourn time under the partial information.

Under total information, an arriving customer observes the number of phases of work in the system, say N, and her expected sojourn time will be equal to

$$\tilde{T}(N) = \frac{N+r}{\mu},$$

for any $N \in 0, 1, 2, \dots$

So, under partial information, an arriving customer observes the number of customers in the system, say k, and her expected sojourn time will be equal to

$$T(k) = \sum_{i=1}^{r} \tilde{T}(r(k-1)+i)\pi(r(k-1)+i \mid k),$$

for any $k \in 0, 1, 2, \dots$, where $\pi(\cdot \mid k)$ is the stationary distribution under the partial information. Therefore, for $i = 1, \dots, r$, $\pi(r(k-1)+i \mid k)$ is the probability of observing $r(k-1)+i$ phases of work, knowing that there are k customers in the system.

Actually, we will see that the stationary distribution depends on the strategy adopted by all customers. So, recalling that σ_n denotes the pure threshold strategy with threshold n, as defined in Remark 3.1, we denote with $\pi_n(\cdot \mid k)$ the stationary distribution conditioned on the partial information under the strategy σ_n, that is when the system has no more than n customers. Similarly we define $T_n(k)$ and $P_n(k)$ as the expected sojourn time and the expected profit function under the strategy σ_n.

Let us now focus on the stationary distribution of the number of customers in the system, whether the strategy adopted is σ_n, $n \geq 3$. The stationary distribution satisfies the *global balance equations*, that are

$$\begin{cases} \pi_n(0)\cdot\lambda = \pi_n(1)\cdot\mu, & \\ \pi_n(N)\cdot(\lambda+\mu) = \pi_n(N+1)\cdot\mu, & N = 1,\dots,r-1 \\ \pi_n(N)\cdot(\lambda+\mu) = \pi_n(N-r)\cdot\lambda + \pi_n(N+1)\cdot\mu, & N = r,\dots,rn-r \\ \pi_n(N)\cdot\mu = \pi_n(N-r)\cdot\lambda + \pi_n(N+1)\cdot\mu, & N = rn-r+1,\dots,rn-1 \\ \pi_n(rn)\cdot\mu = \pi_n(rn-r)\cdot\lambda & \end{cases}, \tag{5.10}$$

which can be rewritten as

$$\begin{cases} \mu\cdot\pi_n(N) = \lambda\cdot\sum_{j=0}^{N-1}\pi_n(j), & N = 1,\dots,r-1 \\ \mu\cdot\pi_n(N) = \lambda\cdot\sum_{j=1}^{r}\pi_n(N-j), & N = r,\dots,rn-r \\ \mu\cdot\pi_n(N) = \lambda\cdot\sum_{j=0}^{rn-N}\pi_n(N-r+j), & N = rn-r+1,\dots,rn \end{cases}. \tag{5.11}$$

Remark 5.4. From the system of equations in (5.11), we observe that, for any $N \in \{0, 1, \dots, rn\}$, $\pi_n(N)$ depends only on values of $\pi_n(\bar{N})$ with $\bar{N} < N$.

Lemma 5.5 (Stationary distribution)**.** The stationary distribution of the number of customers in the system, depending on the threshold strategy σ_n, $n \geq 2$, satisfies the following recursive equation

$$\pi_n(N) = \frac{\pi_n(0)}{\pi_{n-1}(0)}\pi_{n-1}(N), \quad \text{for } N \leq rn - 2r + 1, \tag{5.12}$$

and, for $N > rn - 2r + 1$, $\pi_n(N)$ is defined as in (5.11).

Proof: We prove it by induction. For $n = 2$, we only have to verify that Eq. (5.12) holds for $N = 0, 1$. For $N = 0$, the equation is trivially satisfied, while for $N = 2$, it is satisfied using the first equation of (5.10).

Let $\pi_{n-1}(\cdot)$ be the stationary distribution under the strategy σ_{n-1}, then we prove that $\pi_n(\cdot)$ satisfies Eq. (5.12). From system of equations (5.11), we have that the first $rn - 2r + 1$ equations have the same form both under σ_{n-1} and under σ_n. So, thanks to Remark 5.4, it holds that

$$\frac{\pi_n(N)}{\pi_n(0)} = \frac{\pi_{n-1}(N)}{\pi_{n-1}(0)}, \quad \text{for any } N \leq rn - 2r + 1.$$

□

Therefore, from Lemma 5.5, we have that knowing the stationary distribution $\pi_{n-1}(\cdot)$ allows us to compute the stationary distribution $\pi_n(\cdot)$. Not only that, but it also implies that the conditional stationary distribution $\pi_n(\cdot \mid k)$, conditioned on the partial information of being k customers in the system at an arrival, remains unchanged under the strategy $\sigma_{k+1}, \sigma_{k+2}, \ldots$, while it changes, if it is computed under the strategy σ_k, namely

$$\pi_k(\cdot \mid k) \neq \pi_{k+1}(\cdot \mid k) = \pi_{k+2}(\cdot \mid k) = \ldots$$

As a consequence of this, the same holds for $T_n(k)$ and $P_n(k) = R - CT_n(k)$, that is

$$P_k(k) \neq P_{k+1}(k) = P_{k+2}(k) = \ldots \tag{5.13}$$

It also holds that, for any $k \geq 1$,

$$P_{k-1}(k-1) > P_k(k-1) > P_k(k). \tag{5.14}$$

The second inequality of (5.14) is straight forward, since, when all customers are using the same strategy, entering the system with a larger number of customers in it becomes more expensive in terms of expected sojourn time, that is the expected profit becomes lower. As for the first inequality, however, it is verified because one can prove that $T_{k-1}(k-1) < T_k(k-1)$ by using the equations in (5.11).

Let us define the threshold

$$K = \inf\{k > 0 \mid P_k(k) < 0\}, \tag{5.15}$$

which is well defined since, from (5.14), the function $P_k(k)$ turns out to be decreasing. The following theorem gives the condition to have an equilibrium strategy.

Theorem 5.6 (Equilibrium strategy)**.** Let K be the threshold defined in (5.15). Then σ_K is an equilibrium pure threshold strategy for the $M/E_r/1$ model if and only if $P_K(K-1) \geq 0$.

Proof: Let $F_{K_1,K_2}(k)$ be the payoff of an arriving customer using the strategy σ_{K_1}, when all other customers are using the strategy σ_{K_2}, that is

$$F_{K_1,K_2}(k) = P_{K_2}(k)\mathbb{1}_{\{0\leq k\leq K_1\}}, \quad k = 0, 1, \ldots, K_2.$$

By Definition 1.17, σ_K is an equilibrium strategy if

$$F_{K,K}(\cdot) \geq F_{K_1,K}(\cdot) \quad \text{for any } K_1 > 0. \tag{5.16}$$

If $P_K(K-1) \geq 0$, then the threshold K not only satisfies (5.15), but also

$$K = \inf\{k > 0 \mid P_K(k) < 0\},$$

so, relation (5.16) is satisfied, and σ_K turns out to be an equilibrium strategy.

If, instead, $P_K(K-1) < 0$, then inequality (5.16) is not satisfied because, for example, $P_K(K-1) = F_{K,K}(K-1) < F_{K_1,K}(K-1) = 0$ for $K_1 < K$. So, σ_K cannot be an equilibrium strategy. Not only, also σ_{K_2}, for $K_2 < K$, and σ_{K_3}, for $K_3 > K$ cannot be equilibrium strategies because $0 = F_{K_2,K_2}(K_2) < F_{K_1,K_2}(K_2) = P_{K_2}(K_2)$, for $K_1 > K_2$ and $P_{K_3}(K_3-1) = F_{K_3,K_3}(K_3-1) < F_{K_1,K_3}(K_3-1) = 0$, for $K_1 < K_3$. □

So, in this model, we can find an equilibrium pure threshold strategy only under some conditions. Comparing this model the $M/M/1$ model, we can see that the difference is in the fact that in the $M/E_r/1$ model the expected profit function depends on the strategy adopted by all customers and this affects the possibility of finding equilibrium strategy.

Conclusion

Let us briefly summarize the results obtained in our discussion. After introducing some preliminary results, we studied different queueing networks: the two-node tandem network, the multi-node tandem network and the tree network. In all these models, we assumed that customers arrived according to a Poisson process and were immediately informed about the number of customers in the system, say k. At any queue, service times were independent and exponentially distributed and the service discipline was the FIFO one.

The goal was to find a strategy which turned out to be an equilibrium strategy if all customers adopted it. To do this, we built a profit function $P(k)$, which was proportional to the expected sojourn time $T_i(k)$ at any queue i, and proved that $P(k)$ did not depend on the strategy adopted by all customers. Not only, but we also computed an explicit formula for $P(k)$, proving that it was decreasing and went to $-\infty$ as $k \to +\infty$. These facts allowed us to define the threshold

$$K = \inf\{k \in \mathbb{Z}_+ \mid P(k) < 0\}$$

and prove that the pure threshold strategy with threshold K was the only equilibrium strategy, if $P(K-1) > 0$. If, instead, $P(K-1) = 0$, also the strategy with threshold $K-1$ and any strictly convex combination between these two strategies were equilibria for arriving customers.

In particular, for a tree network, with set of nodes V, the expected sojourn time at queue $i \in V$, knowing the information of being k customers in the system, turned out to be equal to

$$T_i(k) = \frac{\sum\limits_{|n|=k+1} n_i \prod\limits_{\ell \in V} \left(\frac{q_\ell}{\mu_\ell}\right)^{n_\ell}}{\sum\limits_{|n|=k} \prod\limits_{\ell \in V} \left(\frac{q_\ell}{\mu_\ell}\right)^{n_\ell}}, \quad k = 0, 1, 2, \ldots,$$

where, for $\ell \in V$, μ_ℓ was the service rate at queue ℓ and q_ℓ was the probability for a customer to reach queue ℓ from node 1. After this, we defined the expected profit function as

$$P(k) = R - \sum_{i \in V} C_i T_i(k), \quad k = 0, 1, 2, \ldots,$$

where R was the reward a customer got for joining the system and C_i was the unitary cost for sojourn time at queue i, for $i \in V$.

For a multi-node tandem network, with M queues, the same formulas held with $V = \{1, 2, \dots, M\}$ and $q_i = 1$ for any $i \in V$. For a two-node tandem network, however, we used a different approach and the expressions of the expected sojourn times were more simplified and equal to, for $i = 1, 2$,

$$T_i(k) = \begin{cases} \frac{1}{\mu_i - \mu_{3-i}} - \frac{k+1}{\mu_i} \frac{\mu_{3-i}^{k+1}}{\mu_i^{k+1} - \mu_{3-i}^{k+1}}, & \mu_1 \neq \mu_2 \\ \frac{1}{\mu_1}\left(1 + \frac{k}{2}\right), & \mu_1 = \mu_2 \end{cases}.$$

These models are all overtaking free networks, which means that the expected sojourn time of a customer at any queue does not depend on the distribution of customers arrived at the system after her. We proved that, under a FIFO discipline and with exponential service times, a queueing network is overtaking free if and only if it is a tree network. After this, we studied an example of network under the *preemptive-resume label order discipline* and an example of non overtaking free network, proving that there exists an equilibrium strategy in both cases. In conclusion, we showed that for the $M/E_r/1$ model an optimal strategy exists only under some conditions.

Appendix A

MATLAB codes

All MATLAB codes, used for the numerical computations and simulations in the thesis, are loaded in GitHub [19]. We propose, as an example, the function that gives the expected profit of a customer arriving at a tree network and observing k customers in the system.

```
function y=P_tree_network(k,R,v,p,mu,C)

% The function gives the expected profit of a customer entering in a
% tree network, where:
% k is the number of customer in the system that the tagged customer
% observes at her arrival;
% R is the reward that a customer gets to join the system;
% v is a vector that describes the tree: the set of vertices is
% {1,2,3,...,length(v)} and v(i)=j means that node j is parent of
% node i and v(i)=0 means that node i is a root node;
% p is the vector of routing probabilities associated to edges: p(i)
% is the probability associated to edge (v(i),i), if v(i)>0, and
% p(i)=1, if v(i)=0;
% mu is the vector of service rates, where mu(i) is the service rate
% at queue i;
% C is the vector of costs, where C(i) is the cost for unit of sojourn
% time at queue i.

nu=length(v);    % nu is the number of vertices of the tree
q=zeros(1,nu);   % q(i) is the prob. of reaching node i from the root

for i=1:nu
    q(i)=1;
    j=i;
    while v(j) > 0
        q(i)=q(i)*p(j);
        j=v(j);
    end
end
```

```
d=0;
m=matrix(k+1,nu);
for i=1:size(m,2)
    b=C*m(:,i);
    for j=1:nu
        b=b*((q(j)/mu(j))^m(j,i));
    end
    d=d+b;
end

e=0;
mm=matrix(k,nu);
for i=1:size(mm,2)
    a=1;
    for j=1:nu
        a=a*(q(j)/mu(j))^mm(j,i);
    end
    e=e+a;
end

y=R-d/e;

end

% The function matrix creates a matrix, whose columns are all possible
% vectors of length M, whose sum is equal to k.

function y=matrix(k,M)

y=[];
if M==1
    y=k;
else
    for j=0:k
        m=matrix(k-j,M-1);
        a=size(m,2);
        if j == 0
            aa=0;
        else
            mm=matrix(k-j+1,M-1);
            aa=aa+size(mm,2);
        end
        for i=1:a
            y(:,i+aa)=[m(:,i);j];
        end
    end
end

end
```

Bibliography

[1] Ivo Adan and Jacques Resing. Queueing systems. *Department of Mathematics and Computing Science, Eindhoven University of Technology*, 2015.

[2] John Adrian Bondy, Uppaluri Siva Ramachandra Murty, et al. *Graph theory with applications*, volume 290. Macmillan London, 1976.

[3] Pierre Brémaud. *Markov chains: Gibbs fields, Monte Carlo simulation, and queues*, volume 31. Springer Science & Business Media, 2001.

[4] Shelby L Brumelle. On the relation between customer and time averages in queues. *Journal of Applied Probability*, 8(3):508–520, 1971.

[5] Hong Chen and David D Yao. *Fundamentals of queueing networks: Performance, asymptotics, and optimization*, volume 4. Springer, 2001.

[6] Marco Di Summa. Note del corso di ottimizzazione discreta. *Universita degli Studi di Padova*, 2017.

[7] Bernardo D'Auria and Spyridoula Kanta. Pure threshold strategies for a two-node tandem network under partial information. *Operations Research Letters*, 43(5):467–470, 2015.

[8] Agner Krarup Erlang. The theory of probabilities and telephone conversations. *Nyt. Tidsskr. Mat. Ser. B*, 20:33–39, 1909.

[9] Refael Hassin and Moshe Haviv. *To queue or not to queue: Equilibrium behavior in queueing systems*, volume 59. Springer Science & Business Media, 2003.

[10] PW Kanhai. An analysis of the single server queue with erlang and coxian service distributions. *Mathematisch Instituut, Universiteit Leiden*, 2017.

[11] Frank P Kelly and PK Pollett. Sojourn times in closed queueing networks. *Advances in Applied Probability*, 15(3):638–656, 1983.

[12] Bara Kim and Jeongsim Kim. Equilibrium strategies for a tandem network under partial information. *Operations Research Letters*, 44(4):532–534, 2016.

[13] John DC Little. Or forum—little's law as viewed on its 50th anniversary. *Operations research*, 59(3):536–549, 2011.

[14] Daniel Myers. Cs 547 lecture 11: Pasta. *University of Wisconsin-Madison*, 2012.

[15] Pinhas Naor. The regulation of queue size by levying tolls. *Econometrica: journal of the Econometric Society*, pages 15–24, 1969.

[16] Edgar Reich. Waiting times when queues are in tandem. *The annals of mathematical statistics*, 28(3):768–773, 1957.

[17] Edgar Reich. Note on queues in tandem. *The Annals of Mathematical Statistics*, 34(1):338–341, 1963.

[18] Richard Serfozo. *Introduction to stochastic networks*, volume 44. Springer Science & Business Media, 2012.

[19] `https://github.com/Cesi99/queueing-networks.git`.

[20] Jean Walrand and P Varaiya. Sojourn times and the overtaking condition in jacksonian networks. *Advances in Applied Probability*, 12(4):1000–1018, 1980.

[21] Ronald W Wolff. Poisson arrivals see time averages. *Operations research*, 30(2):223–231, 1982.