Operations Research Letters 43 (2015) 467–470

Contents lists available at ScienceDirect

Operations Research Letters

journal homepage: www.elsevier.com/locate/orl

Pure threshold strategies for a two-node tandem network under partial information

Bernardo D'Auria [a,b,*], Spyridoula Kanta [a]

[a] Statistics Department, Madrid University Carlos III, Leganés, Spain

[b] ICMAT, Madrid University Carlos III, Leganés, Spain

ARTICLE INFO

Article history:
Received 13 October 2014
Received in revised form
20 June 2015
Accepted 29 June 2015
Available online 6 July 2015

Keywords:
Tandem network
Individual optimization
Partial information
Threshold strategy

ABSTRACT

In a two node tandem network, customers decide to join or balk by maximizing a given profit function whose costs are proportional to the sojourn time they spend at each queue. Assuming that their choices are taken without knowing the complete state of the system, we show that a pure threshold equilibrium policy exists. In particular we analyze the case when the partial information consists in informing the arrival customers of the total number of users in the network.

1. Introduction

Queueing literature is recently devoting an increasing attention to the economic analysis of queueing systems. Indeed in real applications it is not uncommon that the input to a queueing system is not exogenously defined and is the result of the combined effect of the decisions made by the arriving customers. They may decide whether to join or balk the system according to their convenience and these choices in general lead to a final equilibrium. This phenomenon is mathematically modeled by assuming they are rationally optimizing a given individual profit function. This research, started in the '70s by [8,4], now has reached a good maturity, two central monographs are [6,11]. Most of the literature focuses on a single server system, while we focus here on network models, in particular a series of two $M/M/\cdot$ queues. Previous studies have looked at parallel queues [12,5,7] and for more general topologies extensive studies have been done in the field of telecommunications, see [3,9]. A close model is [1], where a series of queues of $M/M/m$ types is analyzed and the form of the symmetric customer equilibrium is derived together with the explicit socially optimal strategies. The main difference with our model is that there customers make their decisions without getting any information on the state of the system, while here they know the total number of customers already inside. Usually network models show an intrinsic difficulty in getting explicit results, and this partially explains a relatively scarcer literature. The two node tandem network that we study has the advantage of being simpler and allowing a complete analysis. Customers make the decisions to balk or join after knowing how many customers are already in the network. In real applications, it is common that people do not know the complete information on the state of the system, as usually this information is shortly summarized to simplify the decision process. Examples may be found in healthcare systems, where treatment requires two different steps, such as a first queue to get a doctor reservation and a second queue to be attended by the doctor. The interesting result is that the partial information setting simplifies drastically the analysis, allowing to get for this specific case explicit results.

The model is introduced in Section 2, we compute in Section 3 the expected sojourn time of an arriving customer assuming that the full state of the system is known. In Section 4 the same analysis is done when the arriving customers are informed about the total number of customers in the network. Finally we compute the equilibrium strategy in Section 5.

2. The model

We consider a tandem network with two single server nodes with infinite buffers and service times independent and exponentially distributed. Using the index l, with $l = 1$ or 2, to refer to the first or the second node, we denote by μ_l the service

* Corresponding author at: Statistics Department, Madrid University Carlos III, Leganés, Spain.

E-mail address: bdauria@est-econ.uc3m.es (B. D'Auria).

http://dx.doi.org/10.1016/j.orl.2015.06.014
0167-6377/

rate at node l. Customers arrive to the system according to a Poisson process with rate λ and before joining the network they receive partial information about the state of the system. The state space is $\mathbb{N}^2$, that is all possible pairs (Q_1, Q_2) with Q_l the queue length at node l. A *tagged customer* that just arrives, gets a reword R for joining the network, 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 times 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 time, $k = Q_1 + Q_2$, that is

$$P_K(k) = R - C_1 T_{K,1}(k) - C_2 T_{K,2}(k), \tag{1}$$

with $T_{K,l} = \mathbb{E}_K[S_l | Q_1 + Q_2 = k]$. The subindex K tells that the rest of the population is using a pure threshold strategy with threshold K in joining the queue. That is all users besides the tagged one join the network if and only if it contains less than K customers.

The main result of the paper is to show that the tandem network admits a pure threshold K, that is there exists a $K \in \mathbb{N}$ such that

$$P_K(k) \geq 0 \quad \text{as } k < K \quad \text{and} \quad P_K(k) < 0 \quad \text{as } k \geq K.$$

Remark 1. By using the subindex K, we are implicitly assuming that the rest of the population is not allowed to use strategies different from a pure threshold one. This assumption is not restrictive for our purposes, but it does not preclude the existence of policies (even of equilibrium type) that are of a different form.

Remark 2. We always assume that $R > C_1/\mu_1 + C_2/\mu_2$. Being this relation false, a user would get negative net profit even joining an empty network implying a unique equilibrium given by the empty system.

Before analyzing the described model, we first study the case when the complete information is available to the arriving customers. This is done in the next section.

3. Mean sojourn times

Let $S_l(n, m)$ be the sojourn time spent at queue l by a tagged customer that joins a system being in state $(n - 1, m)$, that is she is going to occupy position n in the first queue. Let $T_l(n, m) = E[S_l(n, m)]$ be the corresponding expectation and $T(n, m) = T_1(n, m) + T_2(n, m)$ the total expected sojourn time. The sojourn time in the first queue is Erlang distributed, that is $S_1(n, m) \sim \text{Erlang}(n, \mu_1)$ with mean $T_1(n, m) = n/\mu_1$. The total sojourn time can be computed recursively by applying a first step analysis, that leads to the following formula,

$$\begin{aligned} T(n, m) = {} & \frac{1}{\mu_1 + \mu_2} + \frac{\mu_1}{\mu_1 + \mu_2} T(n - 1, m + 1) \\ & + \frac{\mu_2}{\mu_1 + \mu_2} T(n, m - 1), \quad n, m > 0. \end{aligned} \tag{2}$$

The second term on the right hand side of (2) considers a potential departure from the first queue and the last term a potential departure from the second queue. These events occur with probability $\mu_l/(\mu_1 + \mu_2)$, $l = 1, 2$ respectively. To complete the recursion the following boundary conditions are needed

$$T(0, m) = \frac{m}{\mu_2}; \qquad T(n + 1, 0) = \frac{1}{\mu_1} + T(n, 1), \quad n, m > 0. \tag{3}$$

Using (2) we get a recursive formula to compute $T_2(n, m)$ as shown in the following lemma.

Lemma 3. *The expected sojourn time at the second queue, $T_2(n, m)$, can be computed with the following recursive formula*

$$\begin{aligned} T_2(n, m) = {} & \left(\frac{\mu_2}{\mu_1 + \mu_2}\right)^m T_2(n - 1, 1) \\ & + \frac{\mu_1}{\mu_1 + \mu_2} \sum_{k=0}^{m-1} \left(\frac{\mu_2}{\mu_1 + \mu_2}\right)^k T_2(n - 1, m + 1 - k) \end{aligned} \tag{4}$$

valid for $n > 0$ and $T_2(0, m) = m/\mu_2$, with $m \geq 0$.

Proof. By (2), we get that $T_2(n, m)$ satisfies the following recursive equation

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

and by (3), similar boundary conditions are satisfied. By induction argument it is then straightforward to verify that (4) holds true. □

The following lemma characterizes the conditions under which the T-functions are monotone non-decreasing in the variable n. These conditions are important for the analysis of Section 5.

Lemma 4. *The functions $T_1(n, m)$ and $T(n, k-n)$ are non decreasing in n. The function $T_2(n, m)$ is non decreasing in n if and only if $\mu_1 \geq \mu_2$.*

Proof. The statement is obvious for $T_1(n, m)$ that does not depend on m.

One way to show that the function $T(n, k-n)$ is non decreasing in n, for $n \leq k$ is by proving that $T(n + 1, m) \geq T(n, m + 1)$ by induction using Eqs. (2)–(3). We prefer to use a coupling argument. Using the same probability space, we construct two networks starting respectively with $(n+1, m)$ and $(n, m+1)$ initial users. The proof follows by comparing the waiting times of the customers that are the last ones in the first queue of both networks, and showing that the one in the former network waits more than the corresponding one in the latter. To construct the coupling we assume that the service times for all customers are the same in both networks but we move the customer in service at the first queue of the first network at the end of the queue of the second node of the second network. Since the exit times are ordered by the FIFO discipline and because the moved customer reduces its sojourn time by her service time in the first node, the result holds.

Finally to show that $T_2(n, m)$ is non decreasing in n we prove that $\Delta_1 T_2(0, m) \geq 0$ for all m where $\Delta_1 T_2(n, m) = T_2(n + 1, m) - T_2(n, m)$. From (4), the following holds for any $n > 0$ and $m \geq 0$,

$$\begin{aligned} \Delta_1 T_2(n, m) = {} & \left(\frac{\mu_2}{\mu_1 + \mu_2}\right)^m \Delta_1 T_2(n - 1, 1) + \frac{\mu_1}{\mu_1 + \mu_2} \\ & \times \sum_{k=0}^{m-1} \left(\frac{\mu_2}{\mu_1 + \mu_2}\right)^k \Delta_1 T_2(n - 1, m + 1 - k). \end{aligned} \tag{5}$$

If $\Delta_1 T_2(0, m) \geq 0$ the same holds for $n > 0$ as all the coefficients in (5) are positive. In the opposite case $T_2(n, m)$ is clearly decreasing for some value of (n, m). Let $\alpha = \mu_1/\mu_2$, one can check that

$$\Delta_1 T_2(0, m) = \frac{1}{\mu_2} \left(\frac{\alpha - 1 + (\alpha + 1)^{-m}}{\alpha}\right).$$

The quantity above is decreasing in m. To check that it would be non negative for any value of m we take $m \to \infty$ and get the required condition $\alpha \geq 1$. □