European Journal of Operational Research 237 (2014) 546–554

Contents lists available at ScienceDirect

European Journal of Operational Research

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

Decision Support

Service rate control of closed Jackson networks from game theoretic perspective ☆

Li Xia *

Center for Intelligent and Networked Systems (CFINS), Department of Automation, TNList, Tsinghua University, Beijing 100084, China

ARTICLE INFO

Article history:
Received 1 July 2013
Accepted 20 January 2014
Available online 27 January 2014

Keywords:
Game theory
Queueing
Service rate control
Nash equilibrium
Markov decision processes

ABSTRACT

Game theoretic analysis of queueing systems is an important research direction of queueing theory. In this paper, we study the service rate control problem of closed Jackson networks from a game theoretic perspective. The payoff function consists of a holding cost and an operating cost. Each server optimizes its service rate control strategy to maximize its own average payoff. We formulate this problem as a non-cooperative stochastic game with multiple players. By utilizing the problem structure of closed Jackson networks, we derive a difference equation which quantifies the performance difference under any two different strategies. We prove that no matter what strategies the other servers adopt, the best response of a server is to choose its service rates on the boundary. Thus, we can limit the search of equilibrium strategy profiles from a multidimensional continuous polyhedron to the set of its vertex. We further develop an iterative algorithm to find the Nash equilibrium. Moreover, we derive the social optimum of this problem, which is compared with the equilibrium using the price of anarchy. The bounds of the price of anarchy of this problem are also obtained. Finally, simulation experiments are conducted to demonstrate the main idea of this paper.

1. Introduction

Queueing theory is a well-established methodology in the society of operations research. It can provide a fundamental tool to study the dynamics of many service systems with resource constraints, such as computer systems, communication networks, production systems, and transportation systems. In a queueing system, there widely exists the phenomena of the competition for limited service resources among customers. Thus, the concept of game theory provides a promising research direction for queueing theory. Starting from the pioneering work by Naor in 1969 (Naor, 1969), the game theoretic study of queueing systems attracts considerable research attention in the literature (Altman, Boulogne, El-Azouzi, Jimenez, & Wynter, 2006; Basar & Olsder, 1999; Debo, Parlour, & Rajan, 2012; Guo & Hassin, 2011; Hassin & Haviv, 2003; Xia & Jia, 2013).

According to the categorization standard of game theory, the game theoretic problem in queueing systems is usually a multi-player, stochastic, and non-cooperative game. Moreover, the game theory in queueing systems has some features different from the traditional game theory. First, there are two-level competitions in queueing systems. The first-level competition exists among servers, where servers adjust their strategies to compete for better service profits. The second-level competition exists among customers, where customers compete for more chance to be served. Second, the traditional game theory ignores the networking characteristics of queueing systems, i.e., the servers are interconnected and the customers transit among servers. By utilizing such interconnection structure, it is promising to develop efficient approaches to study the game theory in queueing systems. Perturbation analysis is a successful example and it utilizes the networking characteristics to efficiently estimate the performance gradient or difference of queueing systems (Cao, 1994, 2007; Glasserman, 1991; Gong & Ho, 1987; Ho & Cao, 1991; Leahu, Heidergott, & Hordijk, 2013; Yao & Cassandras, 2012). In this paper, we will study how to utilize the similar idea of perturbation analysis to analyze the game theoretic problem among servers in a closed Jackson network.

Service rate control is a classical optimization problem in queueing theory (Gross, Shortle, Thompson, & Harris, 2008; Stidham, 2011). The goal of service rate control is to identify a set of optimal service rates of all servers to maximize the system average performance. This optimization problem is intensively studied in

☆ This work was supported in part by the National Natural Science Foundation of China (61203039, U1301254), the National 111 International Collaboration Project (B06002), the Specialized Research Fund for the Doctoral Program of Higher Education (20120002120009).

* Tel.: +86 10 62793029.
E-mail address: xial@tsinghua.edu.cn

http://dx.doi.org/10.1016/j.ejor.2014.01.038
0377-2217/© 2014 Elsevier B.V. All rights reserved.

the literature, from the simple queueing systems such as M/M/1 queue or M/M/c queue (George & Harrison, 2001; Neuts, 1978) to complicated queueing networks such as tandem queue, cyclic queue, and Jackson networks (Ma & Cao, 1994; Weber & Stidham, 1987; Xia & Shihada, 2013). Most of the studies of service rate control aim to find the optimal service rates from the perspective of the global system. That is, the optimal service rates correspond to the maximal performance of the entire queueing system. The optimal solution obtained under this scheme is called the *social optimum*.

Much literature of game theory in queueing systems studies the admission control from the customer-level competition, where every customer determines its own strategy to enter the queue or not (Boudali & Economou, 2012; Debo et al., 2012; Guo & Hassin, 2011; Naor, 1969). However, there is little literature about the service rate control of queueing systems from the server-level competition. Only some literature studies the service rate control of a simple queueing system, such as M/M/1 queue (Ata & Shneorson, 2006; Dimitrakopoulos & Burnetas, 2011). This is mainly because the service rate control problem is complicated for queueing networks, such as Jackson networks. Since the game theory is a natural scheme of queueing systems, it is of significance to study the game theoretic control of service rates in queueing networks. In a game theoretic framework of queueing networks, every server optimizes its own service rate control strategy to maximize its own performance (average payoff). Since the servers are interconnected, the performance of a server is affected by not only that server's strategy, but also other servers' strategies. Since the interests of servers are usually conflicting, the servers will compete each other to maximize their own interests. The system will evolve accordingly and it may converge to a *Nash equilibrium*, where every server has no incentive to change its strategy (Nash, 1951).

The above game theoretic control of service rates in queueing networks has practical motivations in many engineering systems. One example is the decentralized power control in wireless networks (Altman et al., 2006; Menache & Ozdaglar, 2011). The communication nodes are viewed as servers and the data packets are viewed as customers. The data packets are transmitted through the network using a multi-hop routing scheme. With the observation of channel status or buffer status, each node has to determine its transmission scheduling strategy to maximize its own payoff (including the power consumption and the throughput). Since the nodes are interconnected through a multi-hop scheme, the scheduling strategies of nodes are mutually affected through the data traffic intensity. This problem is a non-cooperative game. The system may converge to an equilibrium which may be far away from the social optimum. Similar phenomena also exist in other practical problems, such as the intersection traffic control in transportation systems, where each intersection can be viewed as a server and we aim to adjust the green-light period of each intersection to improve the traffic throughput. Therefore, it is meaningful to study the optimization of this game theoretic problem in queueing networks.

In this paper, we study the service rate control problem of closed Jackson networks from a game theoretic viewpoint. Each server is viewed as a player in this game. The payoff of each server includes two parts. One is called the holding cost which is related to the waiting time of customers in the service station. The other is called the operating cost which is the cost to provide certain service rates. Each server controls its own service rates in order to maximize its own average payoff (or minimize the average cost). The control strategies of all the servers are mutually affected through the dynamics of the queueing network. We formulate this problem as an infinite stage non-cooperative stochastic game. We apply the theory of perturbation analysis in Markov decision processes (MDP) (Cao, 2007; Cao & Chen, 1997) to analyze this problem. The theory of perturbation analysis is originally proposed for queueing systems and it can efficiently exploit the interconnection structure of queueing networks to optimize the system performance. For this game theoretic control of service rates in closed Jackson networks, we establish a difference equation which quantifies the difference of average payoffs of each server under any two strategies. We also prove that the average payoff of each server has a monotonic structure with respect to its service rates. No matter what strategies the other servers employ, the best response of a server is to choose its service rates on the boundary. Based on the difference equation, we develop an iterative algorithm to efficiently find the Nash equilibrium under proper conditions. As a comparison, we also study this service rate control problem from a global viewpoint and obtain the social optimum of this problem. The gap between the social optimum and the equilibrium is studied using a metric called the *price of anarchy*. The bounds of the price of anarchy for this problem are also derived. Finally, we conduct numerical experiments to demonstrate the effectiveness of our approach.

The remainder of the paper is organized as follows. In Section 2, we give a formal description of the service rate control problem in closed Jackson networks and formulate it as a non-cooperative stochastic game. In Section 3, we analyze this game theoretic control problem and derive some special properties of this problem. We develop an iterative algorithm to find the equilibrium. The social optimum and the price of anarchy of this problem are also studied. In Section 4, we conduct simulation experiments to demonstrate the main idea of this paper. Finally, we give some discussions and conclude this paper in Section 5.

2. Problem formulation

Consider a closed Jackson network with M servers (Gordon & Newell, 1967; Gross et al., 2008). The total number of customers in the network is a constant N. There is no customer arrival to or exit from the network. The service time of servers is exponentially distributed. The service rate is *load-dependent*, i.e., the server can adjust its service rate according to its queue length. We denote the service rate as μ_{i,n_i}, where n_i is the number of customers at server i, $i = 1,2,\ldots,M$, $n_i = 0,1,\ldots,N$. When $n_i = 0$, $\mu_{i,n_i} = 0$. When a customer joins a server and finds that server busy, this customer will wait in the buffer. The capacity of the buffer is assumed adequate. The service discipline is first come first served. When a customer finishes its service at server i, it leaves server i and joins server j with routing probability q_{ij}, $i,j = 1,2,\ldots,M$. Without loss of generality, we assume $q_{ii} = 0$ for all $i = 1,2,\ldots,M$. Obviously, we have $\sum_{j=1}^{M} q_{ij} = 1$ for all i. The system state is denoted as $\boldsymbol{n} := (n_1, n_2, \ldots, n_M)$. All the possible states compose the state space which is denoted as $\mathcal{S} := \{\text{all } \boldsymbol{n} : \sum_{i=1}^{M} n_i = N\}$.

Each server has its own strategy to determine its own service rates. The value domain of service rate μ_{i,n_i} is denoted as D_{i,n_i}, which is usually a real number interval $[a_{i,n_i}, b_{i,n_i}]$, $i = 1,2,\ldots,M$, $n_i = 1,2,\ldots,N$. Each server has its own payoff function, which consists of two types of costs. One is called the holding cost and the other is called the operating cost. The holding cost reflects the congestion status in the service station. We define the holding cost of server i as $C_h \cdot n_i$ per unit time, where C_h is the holding cost price of each customer in the service station. The operating cost reflects the price of the service rate provided. Higher is the service rate, more is the operating cost. We define the operating cost of server i as $C_o \cdot \mu_{i,n_i}$ per unit time, where C_o is the price to provide a unit service rate. In summary, the payoff function of server i is defined as

$$f_i(\boldsymbol{n}, \boldsymbol{\mu}_{\boldsymbol{n}}) = -C_h \cdot n_i - C_o \cdot \mu_{i,n_i}, \tag{1}$$

where $\boldsymbol{\mu}_{\boldsymbol{n}} := (\mu_{1,n_1}, \mu_{2,n_2}, \ldots, \mu_{M,n_M})$ is the vector consisting of all the service rates of servers at state $\boldsymbol{n}$. Actually, $\boldsymbol{\mu}_{\boldsymbol{n}}$ can also be viewed as