Journal of the Korean Statistical Society 43 (2014) 105–118

Contents lists available at ScienceDirect

Journal of the Korean Statistical Society

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

Approximate queue length distribution of a discriminatory processor sharing queue with impatient customers

Sunggon Kim *

Department of Statistics, University of Seoul, Seoul, 130-743, Republic of Korea

ARTICLE INFO

Article history:
Received 22 December 2012
Accepted 8 April 2013
Available online 29 April 2013

AMS 2000 subject classifications:
primary 60K25
secondary 68M20

Keywords:
Queue length distribution
Impatient customers
Discriminatory processor sharing
Time scale decomposition method

ABSTRACT

We consider a two-class processor sharing queueing system with impatient customers. The system operates under the discriminatory processor sharing (DPS) scheduling. The arrival process of each class customers is the Poisson process and the service requirement of a customer is exponentially distributed. The reneging rate of a customer is a constant. To analyze the performance of the system, we develop a time scale decomposition approach to approximate the joint queue-length distribution of each class customers. Via a numerical experiment, we show that the time scale decomposition approach gives a fairly good approximation of the queue-length distribution and the expected queue length.

1. Introduction

In this paper, we consider a discriminatory processor sharing (DPS) system with impatient customers. DPS is a scheduling algorithm to allocate the service capacity of a system to the multi-class customers. In a DPS system, the service weight is assigned to each class. All of the customers in the system are served simultaneously and a customer's share of the service capacity is proportional to the value of the weight assigned to the customer's class. By giving high value of weights to some classes, the customers of the classes are served preferentially. If all of the customers are served with the same weight, then DPS is reduced to the egalitarian processor sharing (EPS). Recently, EPS and DPS have been studied as a useful abstract model of the flow-level resource sharing in a link of the Internet (Altman, Jimenez, & Kofman, 2004; Massoulie & Roberts, 2001).

Customers are sometimes impatient especially when their waiting time is too long compared to their expectation. In this case, they leave the queue without service completion and find another service station, or come back another time. The situation is the same for online users. Suppose that the response time of a web site is too slow, which may be due to that the web server of the site is too busy (or one of the links of the communication network path from the user to the web site is shared by too many connections) that it takes too much time to load (or transfer) video clips or other types of files. Then the connection to the site may be disconnected and the user tries to find other web sites. Thus, in order to analyze the performance of the multi-class bandwidth-sharing systems with heavy traffic, it is essential to study the queueing behavior of the DPS system with impatient customers.

DPS was proposed by Kleinrock (1967) as an abstraction of the weighted round-robin scheduling for the time-sharing computer system. For the $M/G/1$ DPS queue, Fayolle, Mitrani, and Iasnogorodski (1980) obtained a system of integro-differential equations of the expected sojourn time of each class customers conditioned on the required service time. Rege

* Tel.: +82 222105730.
E-mail address: sgkim@uos.ac.kr.

1226-3192/$ – see front matter © 2013 The Korean Statistical Society. Published by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jkss.2013.04.001

and Sengupta (1994) showed that the conditional sojourn time can be decomposed into a sum of independent random variables. When the service time is exponential, Fayolle et al. (1980) obtained a system of equations of the unconditional expected sojourn time of each class customers, and Kim and Kim (2004) obtained a system of equations of the higher moments of the sojourn time distributions. When the service time follows the phase-type distribution, Haviv and van der Wal (2008) showed that the expected sojourn times satisfy a system of equations. For the moments of the queue-length distributions, Rege and Sengupta (1996) obtained the same result when the service time is exponential. For more results on DPS, we refer the reader to Altman, Avrachenkov, and Ayesta (2006).

To the best of our knowledge, there is no known result on the performance analysis on the DPS system with impatient customers. For the egalitarian processor sharing system with impatient customers (PSI), only a few results are known. The PSI system was first studied by Coffman, Puhalskii, Reiman, and Wright (1994). For the Poisson arrival PSI system with finite buffer, when the service time distribution is exponential and the reneging rate of each customer is constant, i.e. the patience time is exponentially distributed, they obtained a closed form expression of the queue length distribution and the loss rate, and they also obtained the asymptotic behavior of the system as the number of buffers goes to the infinity and also as the ratio of the average time to renege compared to the average service time goes to the infinity. For the Poisson arrival PSI system with heavy tailed service time and general patience time, Guillemin, Robert, and Zwart (2004) obtained the tail asymptotics of the sojourn time distribution in relation to that of the service time distribution. Bonald and Roberts (2003) did an extensive simulation study on the performance analysis of the Poisson arrival PSI system with various service time and patience time distributions. They found that the service time distribution has a significant impact on the performance and the wasted service capacity due to the reneging of the customers increases as the customers become more patient. Ward and Glynn (2003) obtained a diffusion approximation for the queue length process of the Markovian PSI system using a reflected Ornstein–Uhlenbeck process and a reflected affine diffusion process. For the case of $G/G/1$ PSI, Gromoll, Robert, and Zwart (2008) did a fluid limit analysis on the process of a random point measure on $\mathbf{R}_+^2$, where the coordinates of a point denote the residual service time and the residual patience time of a customer in the system. Using the fluid limit, they analyzed the steady-state behavior of the total number of customers in the system.

We consider a two-class $M/M/1$ DPS system with impatient customers. For $i = 1, 2$, the arrival and service rates of the class i customers are λ_i and μ_i, respectively. A customer in the system can leave the system before its service completion at any time, and the patience time is exponentially distributed with rate ν_i for the class i customers, $i = 1, 2$. The capacity of the system is assumed to be C. The input load by the class i customers is $\rho_i = \lambda_i/(C\mu_i)$ while the overall input load is different from ρ_i due to the reneging. The arrived customers are served simultaneously with weights (g_1, g_2), i.e. when there are a number of n_i class i customers, $i = 1, 2$ in the system, the service rate allocated to a class i customer if exists is given by, for $\mathbf{n} = (n_1, n_2)$,

$$r_i(\mathbf{n}) = \frac{g_i}{g_1 n_1 + g_2 n_2} C. \tag{1}$$

It is not easy to analyze the system above exactly. Thus, we adopt the time scale decomposition approach to approximate some performance measures of the system. The time scale decomposition approach is a method developed to study large Markov chains by decomposing a large Markov chain into a number of tractable smaller Markov chains (Yin & Zhang, 2012). Using the time scale decomposition approach, Boxma, Hegde, and Núñez-Queija (2006) and van Kessel, Núñez-Queija, and Borst (2005) analyzed the DPS system where all customers are patient. For an arbitrary number of classes, van Kessel et al. (2005) obtained the approximation of the stationary queue length distribution for the $M/G/1$ DPS system, and Boxma et al. (2006) obtained the stationary queue length and the sojourn time distributions for the two-class $M/M/1$ DPS system with finite buffers. However, their methods are not directly applicable to the DPS system with impatient customers.

In this paper, using the time scale decomposition approach, we obtain some approximate performance measures on the two-class $M/M/1$ DPS system with impatient customers. In Section 2, we introduce how to apply the time scale decomposition approach to our problem. Through the approach, we obtain the approximate joint queue-length distribution of each class customers, the expected queue length, the overall input load, and the throughput, i.e. the average service completion rate of the customers, in Section 3. We analyze the asymptotic behavior of the obtained approximate performance measures as the reneging rate of the customers becomes arbitrarily large in Section 4. The numerical study is given in Section 5 to check how well the obtained results approximate the queue-length distribution and the expected queue length.

2. Time scale decomposition

For $\mathbf{n} = (n_1, n_2)$, we denote by $Q(\mathbf{n})$ the probability that there are n_i class i, $i = 1, 2$, customers in the system at the stationary state. Then, we obtain the following equation:

$$\left(\sum_{i=1}^{2} \lambda_i + (\mu_i r_i(\mathbf{n}) + \nu_i)\, n_i\right) Q(\mathbf{n}) = \sum_{i=1}^{2} \lambda_i (1 - \delta(n_i)) Q(\mathbf{n} - e_i) + (\mu_i r_i(\mathbf{n} + e_i) + \nu_i)(n_i + 1) Q(\mathbf{n} + e_i), \tag{2}$$

where e_i is the unit vector with i'th element of 1, and $\delta(n_i)$ is the Kronecker delta function. It is not easy to derive any tractable solution of $Q(\mathbf{n})$ from the above equation. In the case that the customers of both classes are patient, Rege and