Available online at www.sciencedirect.com
ScienceDirect
European Journal of Operational Research 178 (2007) 463–471
EUROPEAN JOURNAL OF OPERATIONAL RESEARCH
www.elsevier.com/locate/ejor

Stochastics and Statistics

Analysis of a queueing system with a general service scheduling function, with applications to telecommunication network traffic control

Doo Il Choi [a], Tae-Sung Kim [b,*], Sangmin Lee [c]

[a] Department of Applied Mathematics, Halla University, 66 Heungup-li, Heungup-myon, Wonju, Kangwon 220-712, Republic of Korea
[b] Department of Management Information Systems, Chungbuk National University, 12 Gaeshin-dong, Heungduk-gu, Chongju, Chungbuk 361-763, Republic of Korea
[c] Department of Industrial Engineering, Korea Advanced Institute of Science and Technology, Taejon 305-701, Republic of Korea

Received 6 October 2004; accepted 15 December 2005
Available online 17 April 2006

Abstract

In this paper we analyze a queueing system with a general service scheduling function. There are two types of customer with different service requirements. The service order for customers of each type is determined by the service scheduling function $\alpha_k(i,j)$ where $\alpha_k(i,j)$ is the probability for type-k customer to be selected when there are i type-1 and j type-2 customers. This model is motivated by traffic control to support traffic streams with different traffic characteristics in telecommunication networks (in particular, ATM networks). By using the embedded Markov chain and supplementary variable methods, we obtain the queue-length distribution as well as the loss probability and the mean waiting time for each type of customer. We also apply our model to traffic control to support diverse traffics in telecommunication networks. Finally, the performance measures of the existing diverse scheduling policies are compared. We expect to help the system designers select appropriate scheduling policy for their systems.

Keywords: Queueing analysis; Service scheduling function; Traffic control; Telecommunication networks

1. Introduction

Various queueing systems have been studied for traffic control to support traffic streams with different traffic characteristics in telecommunication networks, in particular Asynchronous Transfer Mode (ATM) networks [2,4,5,7,9]. The traffic streams in networks such as voice, data and video have different traffic characteristics. In other words, the voice traffic is delay-sensitive (or real-time) while the data traffic is loss-sensitive (or nonreal-time). Thus, in order to satisfy the quality of service (QoS) of each traffic, a service scheduling

* Corresponding author. Tel.: +82 43 261 3343; fax: +82 43 273 2355.
E-mail address: kimts@chungbuk.ac.kr (T.-S. Kim).

0377-2217/$ - see front matter
doi:10.1016/j.ejor.2005.12.036

policy is necessary. The diverse scheduling methods have been proposed to satisfy the QoS of traffic. The existing representative scheduling methods are the time priority control (for example, the head of the line (HOL) priority scheduling policy) [7] and the loss priority control (for example, partial buffer sharing) [9]. These scheduling methods have focused on satisfying a specific QoS requirement of traffic. For example, the HOL priority scheduling policy is just for the delay-sensitive traffic such as voice. Recently, to satisfy simultaneously QoS of two traffics with different traffic characteristics, a dynamic priority queue has been studied in [2,4,5].

In this paper, we consider a queueing system with a general service scheduling function, which is described as follows: The customers are classified into two types (type-1 and type-2) with different service requirements. There are two buffers I and II to accommodate customers of the type-1 and the type-2 customers respectively. The capacities of buffers I and II are assumed to be finite with size K_1 and K_2 respectively. Arrival of the type-k customers follows a Poisson process with rate λ_k $(k = 1,2)$. The service order for customers of each type is determined by the service scheduling function $\alpha_k(i,j)$ $(k = 1,2)$ where $\alpha_k(i,j)$ is the probability that the type-k customer is selected for service when there are i type-1 customers and j type-2 customers. Clearly, $\alpha_1(i,0) = 1$, $i > 0$, $\alpha_2(0,j) = 1$, $j > 0$ and $\alpha_1(i,j) + \alpha_2(i,j) = 1$ for all i and j. Regardless of the type of customers, the service time is independent and identically distributed with distribution function $G(\cdot)$, mean μ and Laplace transform $G^*(s)$. Even though the service time in applications of ATM networks (the transmission time of a cell (the fixed size of a small packet)) is deterministic, we will consider the service time with a general distribution. The service of customers in each buffer is based on the first-come first-served (FCFS) discipline.

Our queueing model extends the existing scheduling policies and includes these policies as special cases. The important performance measures to satisfy the QoS of each traffic are loss and delay. Thus, we present the loss probability and the mean waiting time by deriving the queue-length distribution. With an application for traffic control to support diverse traffic streams in telecommunication networks, the performance measures of the existing scheduling policies are compared.

We analyze the given queueing system in Section 2. First, we derive the balance equations for the joint queue-length distribution at customer departure epochs by using the embedded Markov chain method. With this information, we express the queue-length distribution at arbitrary times in terms of the distribution at departure epochs, though the latter must be found numerically by solving a finite linear system. Finally we obtain the loss probability and the mean waiting time for each type of customer. In Section 3, some numerical examples are given to compare performance measures such as loss and delay of diverse scheduling policies.

2. Analysis

For analysis of the queueing system, we first use the embedded Markov chain method. The customer's departure epochs are considered as the embedded epochs. To obtain the queue-length distribution at departure epochs, we must know the number of customers arriving during a service time. Thus, let us introduce

$$a_n^r = P\{n \text{ arrivals of type-}r \text{ customer during the service time } S\}$$
$$= \int_0^\infty \frac{e^{-\lambda_r x}(\lambda_r x)^n}{n!}\,dG(x), \quad r = 1,2$$
$$\overline{a}_k^r = \sum_{n=k}^{\infty} a_n^r, \quad r = 1,2.$$

With above notations, we concretely derive the balance equations for the joint queue-length distribution at departure epochs.

2.1. The queue-length distribution at departure epochs

In order to derive the joint queue-length distribution just after departure epochs, let t_n $(n \geqslant 1)$ be the nth departure epoch of a customer with $t_0 = 0$. We also introduce the notations: