대한산업공학회/한국경영과학회 2002 춘계공동학술대회
한국과학기술원(KAIST) 2002년 5월 3일~4일

이산시간 2단계 대기행렬시스템의 분석

Analysis of Discrete-time Two-phase Queueing System

김태성[1] · 장석호[2] · 채경철[2]
[1]충북대학교 경영정보학과/[2]한국과학기술원 산업공학과

Tae-Sung Kim[1] · Seok-Ho Chang[2] · Kyung-Chul Chae[2]

[1]Department of Management Information Systems, Chungbuk National University, Heungduk-gu, Chongju, Chungbuk 361-763, Korea

[2]Department of Industrial Engineering, KAIST, Yousong-gu, Taejon, 305-701, Korea

Abstract

In this paper, we consider a discrete-time two-phase queueing system. We derive the PGF of the system size and show that it is decomposed into two probability generating functions (PGFs), one of which is the PGF of the system size in the standard Geo/G/1 queue. Based on this PGF, we present the performance measure of interest such as the mean number of customers in the system.

Keywords: Discrete-time Queue, Two-phase Queue, Decomposition Property

1. Introduction

Two-phase queueing systems have been discussed in the past due to their applications in various areas such as computer, communication and stochastic systems. In many computer and communication service systems, the situations where arriving packets receive a batch mode service in the first phase followed by individual services in the second phase are naturally common. Recent applications of this queueing system have been discussed by Krishna and Lee [8] and Kim and Chae [7], for example. Note that many papers on two-phase queueing systems have mainly concentrated on the continuous-time models. But there are no results that deals with analysis of the discrete-time two-phase queueing system. As far as we know, this paper appears to be the first one that deals with the discrete-time two-phase queueing system.

In this paper, we consider a discrete-time two-phase queueing system. Packets arrive at the system according to a Bernoulli process and receive batch service in the first phase followed by individual services in the second phase. When the server completes second phase services, if the system becomes empty, it is turned off. After an idle period, if a packet arrives, the server is turned on and begins to serve packets. This type of queueing problem can be easily found in various practical situations:

Consider a central processor connected to a number of peripherals or distributed sub-processors. The central processor collects the jobs arriving at the peripherals or the distributed sub-processors in batches and processes them sequentially. When there is no on-line job for collection, the processor can be switched to process the off-line jobs, to update storage devices or to attend to some maintenance/repair work. As soon as a job arrives, the server is turned on and starts to serve jobs in batch mode.

The rest of the paper is organized as

follows. Section 2 gives assumptions of our system. Section 3 presents the probability generating function (PGF) of the system size and its mean. Finally, Section 4 concludes this paper.

2. System and Assumptions

We assume that the time axis is divided into fixed-length time intervals, called slots and that service times can be started and completed only at slot boundaries and that their durations are integral multiples of slot durations. We will adopt a *Late Arrival System with Delayed Access* (LAS-DA) where packets arrive late during a slot and get delayed access to the server if they arrive to find the system empty. The slot in which a packet arrives is not counted in calculating his/her waiting time. Readers are referred to Bruneel and Kim [2], Hunter [6, p.193] and Takagi [9, p.4] for more details on LAS-DA.

This paper considers the system that satisfies the following assumptions.

Assumptions.
Packets arrive at the system according to a Bernoulli process with mean $1/\lambda$. When a packet arrives, the first phase service starts and the server operates on an exhaustive batch mode service for the packet and the arrivals that occur during the batch service are served. The batch service times $\{B_i,\ i=1,2,\cdots\}$ are independent and identically distributed (*i.i.d.*) random variables with distribution function $B(t)$ and PGF $B(z)$. At the end of the batch service, the entire batch is transferred to the second phase for individual service under the FIFO principle. The individual service times $\{S_i,\ i=1,2,\cdots\}$ are *i.i.d.* with distribution function $S(t)$ and PGF $S(z)$. The arrivals that occur during the individual services have to wait in the first phase for the next batch service. After completing the individual services for the batch in the second phase, the server returns to the first phase to start the next batch service or when the system empties, the server becomes idle. The idle server does not start the batch service until a packet arrives.

The symbols in this paper are defined as follows.

Q: the system size at the beginning of a batch service

Q_1: the system size at the completion of the batch service

Q_2: the system size at the end of the individual services of the batch

B: the batch service time

S: the individual service time

M: the system size at an arbitrary time point

$Q(z), Q_1(z), Q_2(z)$: PGF of Q, Q_1, Q_2

$r(z)$: PGF of M

$B(z), S(z)$: PGF of B, S

$q_0 = P(Q_2 = 0)$

$\gamma = \lambda E(B), \rho = \lambda E(S)$

$b(z) = B(\lambda z + 1 - \lambda)$, $s(z) = S(\lambda z + 1 - \lambda)$

3. Analysis

A. Regeneration cycle analysis

From the definitions of the above variables, the following relations can easily be seen:

Q_1 = Q + the number of arrivals during the batch service

Q_2 = the number of arrivals during the individual services of the Q_1 packets

$$Q = \begin{cases} Q_2 & \text{if } Q_2 > 0 \\ 1 & \text{if } Q_2 = 0 \end{cases}$$

Because the PGFs of the number of arrivals during B and S are $B(\lambda z + 1 - \lambda)$ and $S(\lambda z + 1 - \lambda)$, respectively (see *e.g.* Takagi [9, p.5, (1.9b)]), the above relations are translated to

$$Q_1(z) = Q(z) \cdot B(\lambda z + 1 - \lambda) \qquad (1)$$

$$Q_2(z) = Q_1(S(\lambda z + 1 - \lambda)) \qquad (2)$$

$$Q(z) = [Q_2(z) - q_0] + q_0 z \qquad (3)$$

Combining (1), (2) and (3), we get

$$Q(z)= Q\left(S(\lambda z+1-\lambda)\right)\cdot B\left(\lambda S(\lambda z+1-\lambda)+1-\lambda\right)-q_0(1-z) \quad (4)$$

Differentiating (1),(2) and (3) with respect to z and evaluating at $z=1$ (1), we get

$$E(Q_1)=E(Q)+\gamma \quad (5)$$

$$E(Q_2)=\rho E(Q_1) \quad (6)$$

$$E(Q)=E(Q_2)+q_0 \quad (7)$$

From (5), (6) and (7), we get

$$E(Q_1)=\frac{\gamma+q_0}{1-\rho}$$

Dividing the regeneration cycle into the initial idle period and K sub-service cycles each of which consists of a batch service and the individual services of the batch, we get

$$E(K)=1/q_0$$

Let D be the number of arrivals during the initial delay which consists of a idle period and K first-phase batch service periods in the cycle and Γ be the number of arrivals during the delay-cycle (or, equivalently the regeneration cycle). Then, based on the delay cycle arguments (Takagi [9]), we can find the following expected values.

$$E(D)=1+E(K)\cdot\lambda E(B)=1+\frac{1}{q_0}\gamma \quad (8)$$

$$E(\Gamma)=\frac{E(D)}{1-\rho}=\frac{q_0+\gamma}{1-\rho}\cdot\frac{1}{q_0}=E(Q_1)E(K) \quad (9)$$

Let T_C be the regeneration cycle length. Then the expected value, $E(T_c)$, is obtained using the Wald's equation.

$$E(T_c)=E(\Gamma)E(A) \quad (10)$$

where the expected interarrival time, denoted by E(A), equals λ^{-1}.

Let TP stands for the test packet, then we can derive the following probabilities from (8), (9) and (10), based on renewal reward arguments and on the property of BASTA (Boxma and Groenendijk [1])

Pr(TP arrives during an individual service period)

$$=\frac{E(\Gamma)E(S)}{E(T_c)}=\lambda E(S)=\rho \quad (11.a)$$

Pr(TP arrives during a batch service period)

$$=\frac{E(K)E(B)}{E(T_c)}=\frac{E(B)}{E(Q_1)E(A)}=\frac{\lambda E(B)}{E(Q_1)}=\frac{(1-\rho)\gamma}{\gamma+q_0} \quad (11.b)$$

Pr(TP arrives during an idle period)

$$=\frac{E(A)}{E(T_c)}=\frac{1}{E(Q_1)E(K)}=\frac{q_0}{E(Q_1)}=\frac{(1-\rho)q_0}{\gamma+q_0} \quad (11.c)$$

B. The system size distribution

The key point is that the system under study belongs to the class of Geo/G/1 queues with generalized vacations (see Takagi [9, p.90]) such that both idle periods and first-phase periods act as if they were a vacation period. Thus, the decomposition property of the Geo/G/1 queue with generalized vacations applies to our system. That is, PGF of the system size at an arbitrary slot equals the product of two PGFs; one is the PGF of the system size of an ordinary Geo/G/1 queue (without vacations) at an arbitrary slot, and the other is the PGF of the system size at an arbitrary slot during a vacation period.

The system size PGF for an ordinary Geo/G/1 queue is known as $\frac{(1-\rho)(1-z)s(z)}{s(z)-z}$, where $s(z)=S(\lambda z+1-\lambda)$ (see Takagi [9, p.5 (1.9a)]). To obtain the system size PGF at an arbitrary slot during a vacation period, we use the arrival time approach of Chae et al. [4]. The system size PGF at an arbitrary slot during a vacation period consists of two parts depending on whether TP arrives during an idle period with a (conditional) probability $\frac{q_0}{\gamma+q_0}$, then the PGF equals 1; and if TP arrives during a first-phase period with a (conditional) probability $\frac{\gamma}{\gamma+q_0}$, then the PGF equals $Q(z)\frac{1-b(z)}{\gamma(1-z)}$, where $b(z)=B(\lambda z+1-\lambda)$.

Putting all these together, we finally get the PGF of the system size at an arbitrary slot (or, due to the BASTA property, at TP's arrival point of time).

$$r(z)=\frac{(1-\rho)(1-z)s(z)}{s(z)-z}\left[\frac{q_0}{\gamma+q_0}+\frac{\gamma}{\gamma+q_0}\cdot Q(z)\cdot\frac{1-b(z)}{\gamma(1-z)}\right] \qquad (12)$$

If we differentiate (12) with respect to z and evaluate it at $z=1$, we get the expected system size..

$$E(M)=\rho+\frac{\lambda^2E(S^2)-\lambda\rho}{2(1-\rho)}+\frac{\lambda^2E(B^2)}{2(\gamma+q_0)}+\frac{\gamma}{1-\rho}-\frac{\gamma^2}{\gamma+q_0} \qquad (13)$$

4. Conlusions

In this paper we analyzed a discrete-time two-phase queueing system for the exhaustive batch service. We presented the PGF of the system size and showed that it is decomposed into two PGFs, one of which was the PGF of the system size in the standard Geo/G/1 queue. Based on this PGF, we presented performance measure of interests such as mean number of packet in the system.

Based on the results of this paper, further research is required to study for the discrete-time two-phase queueing system with various threshold policies, such as multiple and single vacations.

References

[1] Boxma, O. J. and W. P. Groenendijk, Waiting times in discrete-time cyclic-service systems, *IEEE Transactions on Communications* 36 (1988), 164-170.

[2] Bruneel, H. and B. Kim, *Discrete-Time Models for Communication Systems Including ATM,* Kluwer Academic Publishers, Boston, MA. 1993.

[3] Bruneel, H. and I. Wuyts, Analysis of discrete-time multiserver queueing models with constant service times, *Operations Research Letters* 15 (1994), 231-236.

[4] Chae, K. C., H. W. Lee, and C. W. Ahn, An arrival time approach to M/G/1-type queues with generalized vacations, *Queueing Systems* 38 (2001), 91-100.

[5] IUT-T, I-Series recommendation on B-ISDN, Geneva, Switzerland, 1994.

[6] Hunter, J. J., *Mathematical Technique of Applied Probability, Vol. II: Discrete-Time Models,* Academic Press, New York, 1983.

[7] Kim, T. S. and K. C. Chae, Two-phase queueing system with generalized vacation, *Journal of Korean Institute of Industrial Engineers* 22 (1996), 95-104.

[8] Krishna, C. M. and Y. H. Lee, A study of two-phase service, *Operations Research Letters* 9 (1990), 91-97.

[9] Takagi, H., *Queueing Analysis, Vol. III: Discrete-Time Systems,* North Holland 1993.