Performance Analysis of a Two-phase Queueing System with Bernoulli Feedback

Doo Il Choi*, Tae-Sung Kim** *Regular Members*

ABSTRACT

We consider a two-phase queueing system with Bernoulli feedback. Customers arrive at the system according to a Poisson process and receive batch service in the first phase followed by individual services in the second phase. Each customer who completes the individual service returns to the tail of the second phase service queue with probability $1-\sigma$. This type of queueing problem can be easily found in computer and telecommunication systems. By deriving a relationship between the generating functions for system size at various embedded epochs, we obtain the system size distribution. The exhaustive and gated cases for the batch service are considered.

I Introduction

We consider a two-phase queueing system with Bernoulli feedback. Customers arrive at the system according to a Poisson process and receive batch service in the first phase followed by individual services in the second phase. Each customer who completes the individual service returns to the tail of the second phase service queue with probability $1-\sigma$. If the system becomes empty at the moment of the completion of the second phase services, the server is idle and must wait until arrival of customer in first queue. The exhaustive and gated cases for the batch service are considered.

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. The processed job may do not meet the predetermined quality standard with probability $1-\sigma$, then the job should be reprocessed. 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. If any job arrives, the server is turned on and starts to serve collected jobs in batch mode.

The queueing system with two-phase service was discussed first by Krishna and Lee[4]. The batch and individual service times in their paper are assumed to be an exponential distribution. Doshi[1] expanded the two-phase queueing system of Krishna and Lee[4] into the case of the batch and individual service times with general distributions. By using an embedded Markov chain approach, Doshi[1] derived the distribution for system size and sojourn time of a customer. Recently, Kim and Chae[2] and Kim and Park[3] studied the two-phase queueing system with vacation and threshold, respectively.

We analyze the two-phase queueing system with Bernoulli feedback. Analysis uses an embedded Markov chain method. That is, by deriving a relationship between generating functions for system size at diverse embedded epochs, we obtain the generating function for system size at a random departure epoch. This information also gives us the mean waiting time of a customer. When $\sigma=1$, our model is just the results of Doshi[1].

* Department of Applied Mathematics, Halla University(dichoi@hit.halla.ac.kr)
** Corresponding author: Department of Management Information Systems, Chungbuk National University(kimts@cbnu.ac.kr)
논문번호: 020420-1004, 접수일자: 2002년 10월 4일

In Section 2, the preliminary for analysis is given In Section 3, we analyze the exhaustive batch service case and obtain the distribution for system size and then the mean waiting time of a customer In Section 4, the gated batch service case is also considered Finally, Section 5 concludes this paper

Ⅱ. Preliminary for analysis

There is a single server, and the first queue Q_1 for the batch service and the second queue Q_2 for the individual service of each customer When the batch service is started, there are two methods: the exhaustive batch service and the gated batch service In the exhaustive batch service, the customers arriving during the batch service are automatically included in the batch service However, in the gated batch service, the batch service includes only customers present at the batch service initiation Therefore, the customers arriving during the batch service have to wait for the next batch service to start. The batch service time is assumed to be independent of their batch size

Introduce the following notations for future analysis.

B= the batch service time
S= the individual service time of each customer

Let F_B be the distribution function of the batch service time (B), F_B^* the LST (Laplace-Stieltjes transform) of B and $\bar{b}$ and $b^{(2)}$ the first two moments of B The individual service times of customers are independent and identically distributed with distribution function F_S, the LST F_S^* and the first two moments $\bar{s}$ and $s^{(2)}$ respectively

Ⅲ Exhaustive batch service case

In this section we consider the two-phase queueing system with Bernoulli feedback when the batch service is exhaustive We first derive the generating functions for number of customers in system at diverse embedded epochs the batch service initiation epoch, the batch service completion epoch (individual service period initiation epoch) and service completion epoch of all customers in the batch (individual service period completion epoch) are considered as the embedded epochs.

Introduce the following notations for n-th cycle of the server:

$N_0(n)$= the number of customers in system at $n-$th batch service period initiation epoch
$N_1(n)$= the number of customers in system at $n-$th batch service period completion epoch
$N_2(n)$= the number of customers in system at $n-$th individual service period completion epoch

$N_0(n)$ is the number of customers in the first queue Q_1 because the second queue Q_2 is empty when a batch service period starts $N_2(n)$ is also the number of customers in Q_1 because Q_2 is empty at the end of an individual service period $N_1(n)$ moves from Q_1 to Q_2 at the completion of the $n-$th batch service period Then, the following relations can be easily seen:

(1) $N_1(n)=N_0(n)+$ {the number of arrivals during the batch service time}

(2) $N_2(n)=$ {the number of arrivals during total individual service time generated by $N_1(n)$ customers}

(3) $N_0(n+1)=\begin{cases} N_2(n) & \text{if } N_2(n)>0 \\ 1 & \text{if } N_2(n)=0 \end{cases}$

Assume $\rho\equiv\bar{\lambda s}/\sigma < 1$ It is necessary and sufficient condition for the stability of the system. Thus, we focus only on the limiting distributions N_0, N_1 and N_2 Each customer who completes individual service rejoins the tail of the second queue with probability $1-\sigma$, $0<\sigma\leq 1$ However, the system size is not changed regardless of the

service order. Thus, we assume that the feedback customers after the individual service completion are served immediately By the fact that the distribution for the served number of customer in Q_2 is given by a geometric distribution starting with one [5], the LST of the distribution function for the length (S_g) of total service of a supercustomer is easily derived by [5]

$$F_g^*(s)=\frac{\sigma F_S^*(s)}{1-(1-\sigma)F_S^*(s)}$$

Let $N_0(z)$, $N_1(z)$ and $N_2(z)$ are the probability generating functions for N_0, N_1 and N_2, respectively. Then, the following probability generating functions are given by (1), (2) and (3)

(4) $N_1(z)=N_0(z)\ F_B^*(\lambda-\lambda z)$

(5) $N_2(z)=\sum_{n=1}^{\infty}E[z^{N_2}\,|N_1=n]\ \Pr[N_1=n]$

$$=\sum_{n=1}^{\infty}[F_g^*(\lambda-\lambda z)]^n\ \Pr[N_1=n]$$

$$=N_1(F_g^*(\lambda-\lambda z))$$

(6) $N_0(z)=N_2(z)-p_{20}(1-z)$,

where $p_{20}=\Pr\{$the system is empty on individual service completion$\}$

Thus, substituting the equations (4) and (5) into (6) we get the following

(7) $N_0(z)=N_0(F_g^*(\lambda-\lambda z))F_B^*(\lambda-\lambda F_g^*(\lambda-\lambda z))$

$$-p_{20}(1-z)$$

To derive $N_0(z)$, let's introduce the following notations

$$L(z)=1-z$$

$$h^0(z)=z,\quad h^{(1)}(z)=h(z)=F_B^*(\lambda-\lambda z)$$

$$g^{(0)}(z)=z,\quad g^{(1)}(z)=g(z)=F_g^*(\lambda-\lambda z)$$

and, for $n\geq 1$,

$$g^{(n)}(z)=g^{(n-1)}(g(z))=g(g^{(n-1)}(z))$$

$$=F_g^*(\lambda-\lambda g^{(n-1)}(z))$$

If $\rho<1$, $z_\infty\equiv 1$ is the unique solution of $z=F_g^*(\lambda-\lambda z)$, $|z|\leq 1$. Since $g^{(n)}(z)$ is analytic and bounded on $|z|\leq 1$, $g^{(n)}(z)$ converges By the fact that $z_\infty\equiv 1$ is the unique solution of $z=F_g^*(\lambda-\lambda z)$,

(8) $g^{(n)}(z)\ \rightarrow\ z_\infty\equiv 1\quad as\ n\rightarrow\infty$

and

(9) $L(g^{(n)}(z))\ \rightarrow\ 0\quad as\ n\rightarrow\infty$

By successive substitution and above results, we obtain

(10) $N_0(z)=\prod_{j=1}^{\infty}h(g^{(j)}(z))$

$$-p_{20}\sum_{k=0}^{\infty}L(g^{(k)}(z))\prod_{j=1}^{k}h(g^{(j)}(z))$$

where $\prod_{j=1}^{0}(\cdot)=1$

Putting $z=0$ in equation (7),

(11) $0=N_0(0)=N_0(F_g^*(\lambda))F_B^*(\lambda-\lambda F_g^*(\lambda))-p_{20}$

because there is at least one customer in system at the batch service initiation epoch.

Thus, by the equations (10) and (11) and the fact $g(0)=F_g^*(\lambda)$, we obtain the following relation

$$p_{20}\ =\ N_0(g(0))h(g(0))$$

$$=\ h(g(0))\Big[\prod_{j=1}^{\infty}h(g^{(j)}(g(0)))$$

$$-p_{20}\sum_{k=0}^{\infty}L(g^{(k)}(g(0)))\prod_{j=1}^{k}h(g^{(j)}(g(0)))\Big]$$

$$=\ \prod_{j=1}^{\infty}h(g^{(j)}(0))-p_{20}\sum_{k=1}^{\infty}L(g^{(k)}(0))\prod_{j=1}^{k}h(g^{(j)}(0))$$

Finally, the p_{20} is given by

(12) $$p_{20}\ =\ \frac{\prod_{j=1}^{\infty}h(g^{(j)}(0))}{1+\sum_{k=1}^{\infty}L(g^{(k)}(0))\prod_{j=1}^{k}h(g^{(j)}(0))}.$$

Next we derive the probability generating function for the number of customers in the system at an arbitrary time Let M be the number of customers present in the system when a tagged customer leaves the system. Also let N_1^t be the size of the batch in which the tagged customer gets moved from the batch service queue to the individual service queue Let K be the position of the tagged customer in that batch Conditioning N_1^t and K in turn, the probability generating function for system size when the tagged customer leaves the system is given by

(13) $M(z)\ \equiv\ E[z^M]$

$$=\ E[E[z^M|N_1^t]]$$

$$=\ \sum_{n=1}^{\infty}E[z^M|N_1^t=n]\Pr\{N_1^t=n\}$$

$$=\ \sum_{n=1}^{\infty}\sum_{k=1}^{n}E[z^M|K=k,N_1^t=n]$$

$$\Pr[K=k|N_1^t=n]\ \Pr\{N_1^t=n\}$$

$$= \sum_{n=1}^{\infty}\sum_{k=1}^{n} z^n \left(\frac{F_g^*(\lambda-\lambda z)}{z}\right)^k \frac{1}{n} \frac{n\Pr\{N_1=n\}}{E[N_1]}$$

$$= \frac{1}{E[N_1]} \frac{F_g^*(\lambda-\lambda z)}{F_g^*(\lambda-\lambda z)-z} \sum_{n=1}^{\infty}[\{F_g^*(\lambda-\lambda z)\}^n - z^n]\Pr\{N_1=n\}$$

$$= \frac{1}{E[N_1]} \frac{F_g^*(\lambda-\lambda z)}{F_g^*(\lambda-\lambda z)-z} [N_1(F_g^*(\lambda-\lambda z)) - N_1(z)]$$

Let $\gamma=\lambda\bar{b}$, from the equations (4) and (7) and

(14) $E[N_1] = \frac{\gamma+p_{20}}{1-\rho}$,

we finally get

(15) $M(z)=\frac{(1-\rho)}{\{\gamma+p_{20}\}} \frac{F_g^*(\lambda-\lambda z)}{(F_g^*(\lambda-\lambda z)-z)}$

$[N_0(z)(1-F_B^*(\lambda-\lambda z))+p_{20}(1-z)]$

By Burke's theorem, the distribution for system size at departures equals to that at arrival instants. Thus, $M(z)$ is the probability generating function for system size at arrival instant Therefore, by PASTA (Poisson Arrivals See Time Average) property, $M(z)$ also is the probability generating function for the number of customers in the system at an arbitrary time. The mean number of customers in the system at an arbitrary time is given by differentiating $M(z)$ at $z=1$:

(16) $E[M]=\rho+\frac{\lambda^2 s_g^{(2)}}{2(1-\rho)}+\frac{1}{2}\frac{p_{20}}{\gamma+p_{20}}$

$+\frac{\rho\gamma+p_{20}}{1-\rho}\frac{\gamma}{\gamma+p_{20}}$

By Little's Law, the mean waiting time of a customer is given by

(17) $E[W] = \frac{E[M]}{\lambda}$.

Remark In our model, the two-phase queue with Bernoulli feedback, it is hard to derive the distribution for waiting time by its computational complexity Thus, we use the little's formula to refer only the mean waiting time.

Ⅳ Gated batch service case

In the gated batch service, only customers in Q_1 present when the batch service is started are served by a batch service Customers arriving after the batch service initiation must wait until the next batch service For the n-th cycle of the server (beginning with the batch service), let's introduce the following notations.

$N_0(n)=$ the number of customers in batch to be served

$N_1(n)=$ the number of customers in Q_1 at batch service completion

$N_2(n)=$ the number of customers in Q_1 when Q_2 becomes empty

We consider the limiting distributions N_0, N_1 and N_2 Then, the following relationships are derived easily

(18) $N_0(z) \equiv E[z^{N_0}] = N_2(z)-p_{20}(1-z)$

(19) $N_1(z,y) \equiv E[z^{N_1}y^{N_0}] = N_0(y)F_B^*(\lambda-\lambda z)$

(20) $N_2(z) \equiv E[z^{N_2}] = N_0(F_g^*(\lambda-\lambda z))F_B^*(\lambda-\lambda z)$

Thus, by the equations (18) and (20), we obtain

(21) $N_0(z)=N_0(F_g^*(\lambda-\lambda z))F_B^*(\lambda-\lambda z)-p_{20}(1-z)$

As in the previous section, $z_\infty\equiv 1$ is the unique solution of $z=F_g^*(\lambda-\lambda z)$, and for any z inside the unit disc, $g^{(n)}(z)\to z_\infty$ as $n\to\infty$. Thus, once again, we use successive substitution to solve (21) for $N_0(z)$ with same notations as in the previous section.

(22) $N_0(z) = \prod_{j=0}^{\infty} h(g^{(j)}(z))$

$-p_{20}\sum_{k=0}^{\infty} L(g^{(k)}(z))\prod_{j=0}^{k-1} h(g^{(j)}(z))$

Since $N_0(0)=0$, by the equation (21) and the fact that $g(0)=F_g^*(\lambda)$ and $h(0)=F_B^*(\lambda)$, $p_{20} = N_0(g(0))h(0)$.

From the equation (22), we thus obtain

(23) $p_{20} = \frac{h(0)\prod_{j=1}^{\infty} h(g^{(j)}(0))}{1+h(0)\sum_{k=1}^{\infty} L(g^{(k)}(0))\prod_{j=1}^{k} h(g^{(j)}(0))}$.

By the same process as in previous section and the equation (21), we obtain the probability generating function for the system size when a randomly tagged customer leaves the system

(24) $M(z) \equiv E[z^M]$

$= \frac{1}{E[N_0]} \frac{F_g^*(\lambda-\lambda z)}{F_g^*(\lambda-\lambda z)-z}[N_0(F_g^*(\lambda-\lambda z))-N_0(z)]$

$= \frac{1-\rho}{\gamma+p_{20}} \frac{F_g^*(\lambda-\lambda z)}{F_g^*(\lambda-\lambda z)-z}$

$[N_0(F_g^*(\lambda-\lambda z))(1-F_B^*(\lambda-\lambda z))+p_{20}(1-z)]$

By Burke's theorem and PASTA property, $M(z)$ also is the probability generating function for the number of customers in the system at an arbitrary time. The mean number of customers in system is given by

$$(25)\quad E[M] = \rho + \frac{1}{2}\frac{\lambda^2 s_g^{(2)}}{1-\rho} + \frac{\rho\gamma + p_{20}}{1-\rho}\frac{\gamma}{\gamma + p_{20}}.$$

V. Conclusions

We analyzed the two phase queueing system with Bernoulli feedback. Although we use the analysis method such as the references [1] and [2], this model with Bernoulli feedback has many applications to computer and telecommunication systems. In particular, our model can be applied to analysis of a central processor connected to a number of peripherals or distributed sub-processors with the probability of not being able to satisfy the predetermined quality standard. Thus, the contribution of our paper is the presentation of two-phase model with Bernoulli feedback, and its explicit derivation of performance measures such as the system size distribution.

References

[1] B. Doshi, "Analysis of a two phase queueing system with general service times", *Operations Research Letters*, Vol. 10, pp. 265-272, 1991.

[2] T. S. Kim and K. C. Chae, "Two-phase queueing system with generalized vacation", *Journal of the Korean Institute of Industrial Engineers*, Vol. 22, pp. 95-104, 1996.

[3] T. S. Kim and H. M. Park, "Cycle analysis of a two-phase queueing model with threshold", *European Journal of Operational Research*, Vol. 144, pp. 157-165, 2003.

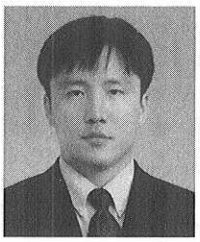
[4] C. M. Krishna and Y. H. Lee, "A study of two-phase service", *Operations Research Letters*, Vol. 9, pp. 91-97, 1990.

[5] H. Takagi, *Queueing Analysis: A Foundation of Performance Evaluation, Vol. 1*, North-Holland, Amsterdam, 1991.

Doo Il Choi

Professor Doo Il Choi received M.S. and Ph.D from Department of applied mathematics of KAIST, and now is in Halla University. His research interest is queueing theory and its applications to telecommunication and computer networks.

Tae-Sung Kim



Professor Tae-Sung Kim has been working for Department of Management Information Systems at Chungbuk National University since September 2000. He received his Bachelor, Master and Doctor degrees in Engineering from Department of Management Science at KAIST in 1991, 1993 and 1997, respectively. He has worked for Internet Economics Team at ETRI(Electronics and Telecommunications Research Institute) as a Senior Researcher from February 1997 to August 2000. His research areas include telecommunications policy, performance evaluation and information security management.