A STUDY OF A QUEUEING SYSTEM WITH THREE-PHASE SERVICE

D. Towsley
Department of Computer and Information Science
University of Massachusetts
Amherst, MA 01003

COINS Technical Report 89-52

June 1, 1989

A Study of a Queueing System with Three-Phase Service

D. Towsley*
Dept. of Computer & Information Science
University of Massachusetts
Amherst, MA 01003

June 1, 1989

Abstract

In this paper we consider a single server that serves customers, each of which requires three phases of service. The server alternates between two queues, a batch queue and an individual queue, giving priority to customers in the latter queue. Customers in the batch queue are served in batches whereas customers in the individual queue are served individually. The batches undergo two service phases. During the first phase, the non-gated phase, any customer entering the batch queue is allowed to join the batch in service. However, once the batch enters the second phase, the gated phase, no other customer is allowed to enter service. At the completion of the gated phase, the customers then enter the individual queue where they acquire their third phase of service. Customers are individually served in the individual queue after which they depart the system. We derive an expression for the mean customer sojourn time under the assumptions of Poisson arrivals, and general service times for each of the three phases.

1 Introduction

In this paper we consider a single server that alternates service between two queues. Customers arrive at one queue called the *batch* queue which is a bulk service queue. After departing the batch queue, customers enter a second FIFO queue lableled the *individual* queue. Whenever

*This work was supported by the Office of Naval Research under Contract N00014-87-K-0796.

there are customers present in both queues, priority is given to customers in the individual queue. Last, customers in the batch queue receive service in batches of size one or more whereas customers are served individually in the individual queue.

Krishna and Lee [2] studied such a system under the assumptions of Poisson arrivals, and exponentially distributed service times for both the batch and individual queues (with different means). They considered two scheduling policies for the batch queue, 1) an exhaustive non-gated policy where all customers present in the queue at the end of a service period are released and 2) an exhaustive gated policy where all customers present at the beginning of service are served. We generalize their analysis to account for generally distributed batch service times and individual service times. We also allow customers to go through two service phases while in the batch queue, a non-gated phase followed by a gated phase. Customers that arrive to the batch queue while the gated phase is in progress are included in the batch that is being served at the time whereas customers that arrive during the gated phase are required to wait until the next non-gated service phase before they can begin service. Last, the the gated phase times are allowed to depend on the number of customers served in a batch. The primary results of this paper are expressions for the mean customer sojourn time and the average number of customers in the system.

A mathematical model of the system is presented in Section 2. The analysis is presented in Section 3 and the results of the paper are summarized in Section 4.

2 The Model

We consider a server that serves two queues, a *batch* queue and an individual queue. When a customer arrives to the system, it enters the batch queue. When it departs the batch queue, it enters the individual queue. After receiving service at the individual queue, the customer departs from the system. The server always serves the individual queue while it is non-empty. When it is empty, the server serves customers in the batch queue. At that time it provides batch service to all of the customers present in the queue at the beginning of service period. The batch service period consists of two phases, the *non-gated phase* followed by the *gated phase*. Customers arriving during the non-gated phase are included in the batch undergoing service and allowed to enter the gated service phase. Upon completion of the gated service phase, the customers enter the individual queue where they individually receive service. We observe that

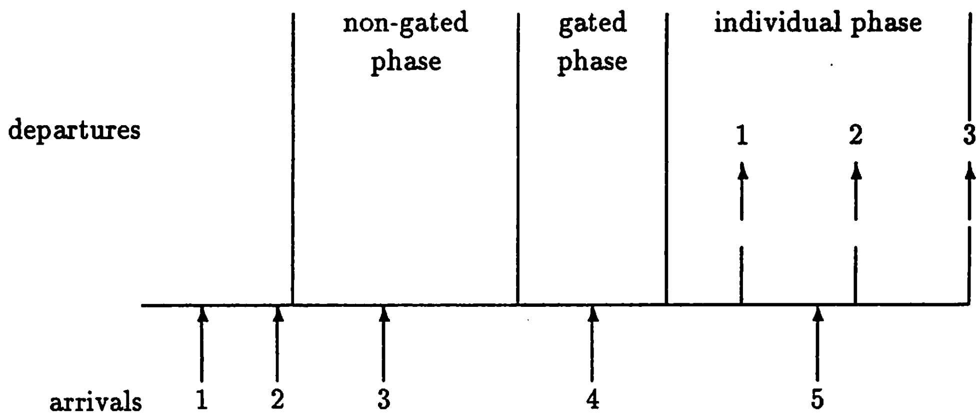

Figure 1: Behavior of the system.

all customers that arrive during the gated phase are required to enter the queue and wait for the next time that the server returns to the queue.

We assume that customers arrive according to a Poisson process with parameter λ, that the non-gated batch phase service times $\{B_k\}_0^\infty$ form an i.i.d. sequence of r.v.'s having probability distribution function $F_B(x) = \Pr[B \leq x]$, $0 \leq x$, and that the individual service times $\{X_i\}_0^\infty$ form an i.i.d. sequence of r.v.'s with distribution function $F_X(x)$, $0 \leq x$. We assume that the i-th gated batch service time can be expressed as $Y_i = R_{i,1} + R_{i,2} + \cdots + R_{i,N_i} + H_i$ where N_i denotes the number of customers in the i-th batch, and $\{H_i\}_0^\infty$ and $\{R_{i,j}, i = 1, N_i\}_0^\infty$ are mutually independent i.i.d. sequences of r.v.'s with distributions $F_H(x) = \Pr[H \leq x]$ and $F_R(x) = \Pr[R \leq x]$.

Time is divided into cycles, each one consisting of an idle period followed by a non-gated batch phase, a gated batch phase and and an individual queue busy period. Let I_k, B_k, Y_k, and Σ_k denote the length of each of these intervals in the k-th cycle. Note that the idle period may be absent, i.e., I_k takes on value 0 with nonzero probability. Figure 1 illustrates the behavior of this system. In this example the idle period is absent. Observe that customer 4 must wait along with customer 5 until the following cycle before receiving service at the batch queue.

We will drop the subscript on the r.v.'s I_k, B_k, Y_k, Σ_k when referring to the stationary versions of those r.v's. Let Z denote a continuous valued r.v. We will adopt the convention that $\overline{Z} = E[Z]$, $\overline{Z^2} = E[Z^2]$ and $Z^*(s) = E[e^{-sZ}]$. In the case that Z is a discrete valued r.v., we will adopt the same convention for the moments and $G_Z(z) = E[z^Z]$.

3 Analysis

Let W denote the stationary time that a customer spends in the system until it enters the gated batch phase. Let T denote the stationary customer sojourn time, i.e., time until it leaves the system. Last, let S denote the time that the customer spends in the individual queue. We have $T = W + Y + S$ and as a consequence,

$$\overline{T} = \overline{W} + \overline{Y} + \overline{S}. \tag{1}$$

We focus on $\overline{W}$ first. Consider a randomly chosen customer, J, and define the following events.

- E_0 - the event that J finds the server idle at arrival,
- E_1 - the event that J finds the server in the non-gated batch phase at arrival,
- E_2 - the event that J finds the server in the gated batch phase at arrival,
- E_2 - the event that J finds the server at the individual queue at arrival.

We condition W on each of these events. The conditional expectations of W conditioned on each of these events are

$$E[W|E_0] = \overline{B}, \tag{2}$$

$$E[W|E_1] = \overline{B^2}/\left(2\overline{B}\right), \tag{3}$$

$$E[W|E_2] = \overline{Y^2}/\left(2\overline{Y}\right) + \overline{\Sigma} + \overline{B}, \tag{4}$$

$$E[W|E_3] = \overline{\Sigma^2}/\left(2\overline{\Sigma}\right) + \overline{B}. \tag{5}$$

Consider the gated batch phase time. We have expressed Y as a sum of H and a random number of R's. Hence results from [1, pp. 111-112] allows us to write

$$\overline{Y} = \overline{H} + \overline{R}\,\overline{N}, \tag{6}$$

$$\overline{Y^2} = \overline{N^2}\,\overline{R}^2 + (\overline{R^2} - \overline{R})\overline{N} + \overline{H^2} + 2\overline{H}\,\overline{R}\,\overline{N}. \tag{7}$$

Consider the individual queue busy period, Σ. We can express Σ as the sum of N individual service times. Hence, from [1, pp. 111-112]

$$\overline{\Sigma} = \overline{N}\,\overline{X}, \tag{8}$$

$$\overline{\Sigma^2} = (\overline{N^2} - \overline{N})\overline{X}^2 + \overline{N}\,\overline{X^2}. \tag{9}$$

We focus now on the probabilities of the events E_0, E_1, E_2, E_3. Let C denote the length of a cycle, $C = I + B + Y + \Sigma$. The probabilities of the four events are

$$\Pr[E_0] = \overline{I}/\overline{C}, \tag{10}$$

$$\Pr[E_1] = \overline{B}/\overline{C}, \tag{11}$$

$$\Pr[E_2] = (\overline{N}\,\overline{R} + \overline{H})/\overline{C}, \tag{12}$$

$$\Pr[E_3] = \overline{\Sigma}/\overline{C}. \tag{13}$$

We know the statistics of B and we have expressed the statistics of Σ in terms of the statistics of N; hence we focus on I. Let A denote the number of customers that arrive during a gated batch phase and an individual busy period. The event $I > 0$ corresponds to the event that $A = 0$. Hence we write

$$\overline{I} = \Pr[A = 0]/\lambda$$

Because of the assumption that customers arrive according to a Poisson process, we have $G_A(z) = G_N(X^*(\lambda(1-z))R^*(\lambda(1-z)))$. Consequently $\Pr[A = 0] = G_A(0) = G_N(D^*(\lambda))$ and

$$E[I] = G_N(D^*(\lambda))/\lambda,$$

$$E[C] = G_N(D^*(\lambda)) + \overline{B} + \overline{N}(\overline{X} + \overline{R}) + \overline{H}$$

where $D^*(s) = X^*(s)R^*(s)$.

We now remove the conditioning on W to obtain

$$\overline{W} = \frac{p_0\overline{B}/\lambda + \overline{B^2}/2 + \overline{\Sigma^2}/2 + \overline{Y^2}/2 + \overline{\Sigma}(\overline{Y} + \overline{B}) + \overline{Y}\,\overline{B}}{p_0/\lambda + \overline{B} + \overline{N}(\overline{X} + \overline{R}) + \overline{H}}. \tag{14}$$

where

$$p_0 = \Pr[A = 0] = G_N(D^*(\lambda)). \tag{15}$$

Consider the individual queue sojourn time, S. J arrives as part of a batch of jobs and waits for all customers ahead of it to be served. The nunber of customers ahead of J corresponds to the residual number of customers in a batch that depart the batch queue. Hence, we have

$$\overline{S} = \left[\frac{\overline{N^2}}{2\overline{N}} + \frac{1}{2}\right]\overline{X}. \tag{16}$$

We are left with the task of determining the statistics of N.

Theorem 1 *The p.g.f. $G_N(z)$ satisfies the following functional equation*

$$G_N(z) = [G_N(D^*(\lambda(1-z)))H^*(\lambda(1-z)) + (z-1)p_0 H^*(\lambda)]\,B^*(\lambda(1-z)). \tag{17}$$

where p_0 is given by the expression in equation (15).

Proof. Let N_i denote the number of customers served during the i-th cycle. Observe that this corresponds to the number of customers that arrive during the gated and individual service phases of the $(i-1)$-th cycle and the non-gated phase of the i-th cycle. Hence $\{N_i\}_{i=0}^{\infty}$ is a Markov chain. The chain is ergodic whenever $\lambda(\overline{X} + \overline{R}) < 1$. Let π_j, $j = 1, 2, \cdots$ denote the stationary distribution for this Markov chain. These probabilities satisfy

$$\pi_i = \sum_{j=0}^{\infty} \pi_j q_{j,i}, \quad i = 0, 1, \cdots \tag{18}$$

where

$$\begin{aligned} q_{j,i} &= \int_0^{\infty} (1+\lambda x)e^{-\lambda x} f_\tau(x,j)dx \int_0^{\infty} \frac{(\lambda y)^{i-1}}{(i-1)!}e^{-\lambda y} f_B(y)dy \\ &+ \sum_{l=2}^{i} \int_0^{\infty} \frac{(\lambda x)^l}{l!}e^{-\lambda x} f_\tau(x,j)dx \int_0^{\infty} \frac{(\lambda y)^{i-l}}{(i-l)!}e^{-\lambda y} f_B(y)dy, \quad i = 1, 2, \cdots \end{aligned}$$

Here τ denotes the length of the gated and individual service phases, $\tau = H + \sum_{l=1}^{N}(X_l + R_l)$ and $f_\tau(x,j)$ is the probability density function of τ conditioned on $N = j$. Letting $D_l = X_l + R_l$ we have $E[e^{-\tau s}|N = j] = H^*(s)D^*(s)^j$, $j = 1, \cdots$. Multiplying both sides of equation (18) by

z^i and summing over i yields

$$\begin{aligned}
G_N(z) &= \sum_{j=1}^{\infty}\pi_j\sum_{l=1}^{i}\int_0^\infty \frac{(\lambda x z)^l}{l!}e^{-\lambda x}f_\tau(x,j)dx\int_0^\infty \frac{(\lambda y z)^{i-l}}{(i-l)!}e^{-\lambda y}f_B(y)dy \\
&\quad + \sum_{j=1}^{\infty}\pi_j\int_0^\infty z e^{-\lambda x}f_\tau(x,j)dx\int_0^\infty \frac{(\lambda y z)^{i-1}}{(i-1)!}e^{-\lambda y}f_B(y)dy \\
&= H^*(\lambda(1-z))B^*(\lambda(1-z))\sum_{j=1}^{\infty}\pi_j D^*(\lambda(1-z))^j \\
&\quad + (z-1)B^*(\lambda(1-z))H^*(\lambda)\sum_{j=1}^{\infty}\pi_j D^*(\lambda)^j \\
&= H^*(\lambda(1-z))B^*(\lambda(1-z))G_N(D^*(\lambda(1-z))) \\
&\quad + (z-1)B^*(\lambda(1-z))H^*(\lambda)G_N(D^*(\lambda))
\end{aligned}$$

which yields equation (17). □

We define the function $g : [0,1] \to [0,1]$,

$$g(z) = D^*((\lambda(1-z)), \quad 0 \le z \le 1 \tag{19}$$

and the notation $g_i(x)$ to denote the i-th fold composition of g, i.e.,

$$g_i(x) = \begin{cases} z, & i = 0, \\ g(g_{i-1}(z)), & i = 1, 2, \cdots. \end{cases} \tag{20}$$

The p.g.f. $G_N(z)$ can be expressed as

$$G_N(z) = (N_1(z) - p_0 H^*(\lambda)N_2(z))B^*(\lambda(1-z)) \tag{21}$$

where

$$\begin{aligned}
N_1(z) &= \prod_{i=0}^{\infty} H^*(\lambda(1-g_i(z)))B^*(\lambda(1-g_{i+1}(z))), \\
N_2(z) &= (1-z) + \sum_{i=1}^{\infty} d_i(z)(1-g_i(z)), \\
d_i(z) &= \prod_{j=0}^{i-1} H^*(\lambda(1-g_i(z)))B^*(\lambda(1-D^*(\lambda(1-g_i(z))))).
\end{aligned}$$

This is verified by substituting (21) into (17). The unknown constant, $p_0 = G_N(D^*(\lambda))$ is obtained from the above expression to be

$$p_0 = \frac{N_1(D^*(\lambda))B^*(D^*(\lambda))}{1 + H^*(\lambda)B^*(D^*(\lambda))N_2(D^*(\lambda))}. \tag{22}$$

The expression for $G_N(z)$ is well behaved in $[0,1]$ when $\lambda(\overline{X} + \overline{R}) < 1$. This is because g can be shown to be a contraction mapping from $[0,1]$ into itself with a unique fixed point at $z = 1$ when $\lambda(\overline{X} + \overline{R}) < 1$. The evaluation of p_0 requires a numerical calculation of $N_1(D^*(\lambda))$ and $N_2(D^*(\lambda))$ which are expressed as an infinite sum and infinite product respectively. In practice, these are approximated by a finite sum and product. If we use the superscript (n) to denote such approximations based on the first n terms, then the following expressions provide error bounds on the truncated values of $N_1()$ and $N_2()$ respectively

$$\begin{aligned} N_2(D^*(\lambda)) - N_2^{(n)}(D^*(\lambda) &\leq \rho^n(1 - D^*(\lambda)/(1-\rho), \;\; n = 0, 1, \cdots, \\ N_1(D^*(\lambda))/N_1^{(n)}(D^*(\lambda)) &\leq \exp(\rho^{n+1}/(1-\rho)[\overline{H}/(1-\rho^n\overline{H}) + \rho\overline{B}/(1-\rho^{n+1}\overline{B})], \\ & \qquad n \geq \min\{0, -\ln\overline{H}/\ln\rho, -\ln/\rho - 1\} \end{aligned}$$

where $\rho = \lambda(\overline{X} + \overline{R})$.

By differentating both sides of equation (17) and making use of the moment generating properties of the p.g.f., we obtain

$$\overline{N} = \frac{\lambda(\overline{H} + \overline{B}) + p_0 H^*(\lambda)}{1 - \rho}, \tag{23}$$

$$\overline{N^2} = \frac{\overline{N}[1 + 2\lambda^2(\overline{X} + \overline{R})(\overline{B} + \overline{H}) + \lambda^2(\overline{X^2} + \overline{R^2} - \overline{X}^2 - \overline{R}^2)]}{1 - \rho^2} \tag{24}$$

$$+ \frac{\overline{B^2} + 2\lambda^2\overline{B}\,\overline{H} + 2\lambda\overline{B}p_0 H^*(\lambda) + \lambda^2\overline{H^2}}{1 - \rho^2}. \tag{25}$$

Substitution of the expressions for $\overline{N}$ and $\overline{N^2}$ into equations (6), (14), (16), and (1) yields $\overline{T}$. It is possible to obtain expressions for the p.g.f. for the distribution of the total number of customers in the system and the Laplace transform for the sojourn time of a randomly chosen customer using similar calculations. In this case, care must be taken to capture the dependencies between the statistics of I, Y, and Σ.

In the remainder of this section, we adapt the above results to the two models considered in [2].

Exhaustive non-gated service: In this sytem, the gated phase is nonexistant, i.e., $Y = 0$. The mean customer sojourn time is

$$\overline{T} = \frac{p_0 \overline{B}/\lambda + \overline{B^2}/2 + \overline{\Sigma^2}/2}{p_0/\lambda + \overline{B} + \overline{N}\overline{X}} + \frac{\overline{N^2}\,\overline{X}}{2\overline{N}} + \frac{\overline{X}}{2}$$

where

$$\begin{aligned}
\overline{N} &= \frac{\lambda \overline{B} + p_0}{1-\rho}, \\
\overline{N^2} &= \frac{\overline{N}[1 + 2\lambda\rho\overline{B} + \lambda^2(\overline{X^2} - \overline{X}^2)] + \overline{B^2} + 2\lambda\overline{B}p_0}{1-\rho^2}, \\
p_0 &= G_N(X^*(\lambda)), \\
\rho &= \lambda\overline{X}.
\end{aligned}$$

The quantities $\overline{\Sigma}$ and $\overline{\Sigma^2}$ are given by equations (8) and (9).

Exhaustive gated service: In this sytem, the non-gated phase is nonexistant, i.e., $B = 0$. In addition, the gated phase service time is independent of the number of customers served. he mean customer sojourn time is

$$\overline{T} = \frac{(\overline{\Sigma^2} + \overline{H^2})/2 + \overline{\Sigma}\,\hbar}{p_0/\lambda + \overline{B} + \overline{N}\overline{X}} + \overline{Y} + \frac{\overline{N^2}\,\overline{X}}{2\overline{N}} + \frac{\overline{X}}{2}$$

where

$$\begin{aligned}
\overline{N} &= \frac{\lambda \overline{H^2} + p_0 H^*(\lambda)}{1-\rho}, \\
\overline{N^2} &= \frac{\overline{N}[1 + 2\lambda\rho\hbar + \lambda^2(\overline{X^2} - \overline{X}^2)] + \lambda^2\hbar^2}{1-\rho^2}, \\
p_0 &= G_N(X^*(\lambda)), \\
\rho &= \lambda\overline{X}.
\end{aligned}$$

The quantities $\overline{\Sigma}$ and $\overline{\Sigma^2}$ are given by equations (8) and (9).

4 Summary

In this paper we have presented and analyzed a model of a system in which customers receive three phases of service. This work generalizes the results of Krishna and Lee for two phase service systems. Their two models are obtained by restricting the service times to be exponential r.v.'s and by setting the appropriate service phase to be of zero duration.

References

[1] A. Drake, *Fundamentals of Applied Probability*, Mc-Graw Hill, Inc., New York, 1967.

[2] C.M. Krishna, Y.H. Lee, "A Study of Two-Phase Service", submitted to *Operations Research Letters*.