APPROXIMATE ANALYSIS FOR BULK QUEUEING SYSTEM
WITH COMPOSITE SERVICE DISCIPLINE

Kazuhiro Ohtsuki*, Toshio Shimizu** and Toshiharu Hasegawa*

*: Dept. of Appl. Math. & Physics, Kyoto University, Japan
**: Application System Development Section
Matsushita Electric Industrial Co., LTD., Japan

ABSTRACT

In this paper, we consider a transportation type bulk arrival bulk service queueing system with composite bulk service discipline as a fundamental research on future communication networks. In this system, customers arrive in group and customers in a group have an identical attribute. From the head of the queue, customers with the same attribute are served within a finite bulk size at the same time. For such bulk service queueing system, the average queue length of customers in the system is approximately analyzed. In our approach, each class of customers with the same attribute is assumed to form their own queue. An approximation method for obtaining queue length of each queue just before the service occasion epoch, where probability generating function approach and embedded Markov chain method are utilized. Numerical examples show our approximation results are well verified by the simulation ones.

1. INTRODUCTION

This paper considers a queueing system where customers arrive in group according to compound Poisson process and form a single queue. Customers in a group have an identical attribute. Waiting customers are served by a single server in a bulk service fashion according to the predetermined discipline. From the head of the queue, customers with the same attribute are served within a finite bulk size at the same time. Customers in a group may not be served simultaneously, then a group is divided into two or more subgroups. Customers having arrived in different groups can be served in the same bulk if they have an identical attribute. A service is carried out during a time interval between service occasion epochs. The time intervals between service occasion epochs are independently and identically distributed with an arbitrary distribution.

This kind of queueing system can be applied to analyze some of stochastic behaviors in communication system as a mathematical model. Examples are following. In a packet switching network, each switching node assembles several packets having the same attribute as a common destination into a frame with fixed length. And gate nodes perform almost the same role to connect multiple local area networks through terrestrial global network or satellite communication network, and in the case of ISDN, all external interface will play the same role. Even in satellite communication system terrestrial stations can be considered to have a similar discipline to increase the utilization factor in the future. In those cases, a customer, an attribute, a server and a service time are regarded as a packet, a destination or a kind of packet, a switching equipment for framing and a time interval of sending a frame, respectively. And a service occasion epoch is regarded as a time epoch that a gate opens, or at a time boundary point of a slot in a synchronized system (in this case, the time intervals of service occasion epochs are assumed to obey a constant distribution).

The already published works for bulk service queueing systems are almost concerned with the systems where all customers have an identical attribute, called HBQS(homogeneous bulk queueing system). Bailey [1] derived the equilibrium distribution of queue length in a HBQS, where customers arrive individually, by the embedded Markov chain method. Jaiswal [4] solved the same problem as the one of Bailey except that the maximum number of customers to be served at the same time is not constant and obtained time-dependent solution by phase method [5]. Miller [6] studied group arrival and group service, namely M/G/1 queue. Many researchers (Bhat [2] Cohen [3] Neuts [7] et al,) considered various versions of HBQS. Watanabe et al. [8] proposed Exclusive Group Service (EGS) discipline in M/G/1 type bulk queueing system, where arriving groups have different attributes though customers in group have an identical attribute. Our model proposed here has a composite bulk service discipline which allows different groups to have the same attribute.

For such bulk service queueing system, the average queue length of customers in the system is approximately analyzed. In our approach, each class of customers with the same attribute is assumed to form their own queue. There exist as many queues as the number of different attributes. An approximation method for obtaining queue length of each queue just before the service occasion epoch, where probability generating function (p.g.f.) approach and embedded Markov chain method are utilized. Our approximation results are well verified by the simulation results. The simulation results do not have tendency to underestimate the average number of customers.
2. MODEL

Let us consider a transportation type bulk arrival bulk service queueing system model under composite bulk service discipline.

Customers arrive in groups according to compound Poisson process, that is, the groups of customers arrive according to Poisson process and the sizes of groups, the number of customers in a group, are independent and identically distributed with an arbitrary distribution. Furthermore all customers of a group have the same attribute. The service, which is in batch of fixed capacity, begins just after a service occasion epoch and the time intervals of service occasion epochs are independently and identically distributed with an arbitrary distribution. The server is able to accommodate as many customers as possible within the capacity if at the back there are the groups of customers whose attributes are the same as the first one in the queue. If there are no customers in the queue at the time of completing the service, next service does not start as soon as the customers arrive at the system but starts after a next service occasion epoch.

In Fig.1, a symbol stands for a customer and the same symbol means an identical attribute. The number figured in a symbol is the group number arriving the node. Fig.1-(a) shows a situation just before service starting. The queue consists of four groups and the number of customers of the head group is 2. Customers of the head group in the queue have a claim to be served at first within the bulk size which is 4 in this figure. If the server can serve more customers yet, he accommodates the subsequent customers who have the same attribute as the head group. Then, the system state changes as Fig.1-(b). In our approach, each class of customers with the same attribute is assumed to form their own queue (see Fig.2). There exist as many queues as the number of different attributes. The service discipline is put another way as follows. Customers in the queue in which the earliest arrival group have joined are served within the capacity according to FCFS discipline. Note that FCFS discipline only applies to the customers (or groups) who have the same attribute. For example, in the case of Fig.2, group 4 is served earlier than group 2 and 3, although he arrived later.

3. ANALYSIS

3.1. Approximate Analysis for Mean Queue Length

We propose approximation method of obtaining a probability generating function for queue length of customers immediately before the beginning of service under a composite bulk service discipline. The outline of our method is as follows. We pay attention to the behavior of customers who have the same attribute, that is, we analyse the average length of tagged queue under the assumption that if any, are served is obtained. Let us introduce the following notations (random variables) to formulate our queueing model. We assume that queue i is formed by only customers whose attribute is i and has an infinite buffer.
The time interval of the j-th service occasion epoch and j+1-st one, we call $V_j$ as j-th time interval.

The number of arriving customers at queue i during the time interval $V_j$, we call $X_j^i$.

The number of waiting customers in queue i immediately before the j-th service occasion epoch, we call $W_j^i$.

The number of customers in queue i who are accommodated during the j-th time interval, we call $Y_j^i$.

The size of k-th arriving customer-group in queue i. Let us call $B_k^i$ as the customer-group size in queue i.

In what follows, we pay attention to queue 1 (i.e. customers of attribute 1) and analyse the queue length of queue 1. To make simple, the above random variables are rewritten as follows.

$X_j = X_j^1$, $W_j = W_j^1$, $Y_j = Y_j^1$, $B_k = B_k^1$.

Accordingly, the following recurrence relations can be provided for the queue length $W_j$ of customers and the number of served customer $Y_j$, as shown in Fig. 3.

$$W_{j+1} = W_j - Y_j + X_j, \quad (1)$$

$$Y_j = \begin{cases} \min\{W_j, c\} & \text{(the case where customers can be served)} \\ 0 & \text{(otherwise).} \end{cases} \quad (2)$$

Eq. (2) shows that the number of served customers, $Y_j$, depends on whether customers in queue 1 can be served or not rather than the queue length $W_j$. Therefore, it is necessary to grasp not only the situation of queue 1 but also that of other queue. We must know which head customers of each queue arrive earliest of all in order to analyze this system exactly. Thus, an exact analysis is extremely difficult for the queueing model like this. Then approximation method for queue length $W_j$ are proposed by using the probability that customers in queue 1 are able to be collected.

If subscript j of the above random variables is omitted, it means that j need not be considered.

First, the probability $p_k$ that k customers arrive at queue 1 during a current service is,

$$p_k = \Pr[X_j=k] = \Pr[X=k] = \sum_{n=0}^{\infty} \Pr[X=k; V=t] dV(t) \quad (3)$$

where $b_k = \Pr[B_1+B_2+\cdots+B_n=k]$.

The probability generating function, p.g.f., $P(z)$ of $X$ is derived from both the p.g.f. $B(z)$ of $B$, i.e., the size of a customer group at its arrival time, and $V(t)$.

$$P(z) = \sum_{k=0}^{\infty} p_k z^k \quad (4)$$

$$B(z) = \sum_{k=0}^{\infty} b_k z^k \quad (5)$$

According to Eqs (1) and (3), the following equation is held.

$$\Pr[W_{j+1}=k | W_j=n, Y_j=m] = \frac{p_k}{P} = q \Pr[W_j=n] \tag{6}$$

$$\Pr[W_{j+1}=k | W_j=n, Y_j=0] = (1-q) \Pr[W_j=n] \tag{7}$$

Multiplying by $z^k$ on both sides of the above equation and adding over k, we have:

$$\frac{P}{\Pr[W_j=n]} = \sum_{n=0}^{\infty} \Pr[W_j=n] \tag{8}$$

Where, $SV(j)$ and $NSV(j)$ denote the event that the customers in queue 1 can be served or not during the j-th service occasion epoch, $q$, is independent of the queue length $W_j$, and identical. The approximation method for $q$ is explained in Section 3.2.

$$\Pr[W_j=n, SV(j)] = q \Pr[W_j=n] \tag{9}$$

$$\Pr[W_j=n, NSV(j)] = (1-q) \Pr[W_j=n] \tag{10}$$

3.2A-23
Introducing Eqs.(9) and (10), Eq.(8) is rewritten as follows.

\[ R(z) = \sum_{n=0}^{\infty} \Pr[W_{j+1}=k]z^n \]

where,

\[ A(z) = q \sum_{n=0}^{c-1} w_n(c-c-n) \]

and

\[ D(z) = z^2(F^1(z)-(1-q))q \]

The \( c \) unknown variables \( w_n \) \((n=0,1,...,c-1)\) can be determined from the facts that \( R(z) \) is analytical in \(|z|<1\) and that \( D(z)=0 \) has the \( c \) zeroes in \(|z|<1\) by Rouche's theorem. If we assume that \( D(z)=0 \) has \( 1+2,3,...,c \) as its simple zeroes, \( w_n \) \((n=0,1,...,c-1)\) are obtained as the solution of the following linear equations.

\[ A(1) = D(1) \]

\[ c-1 \sum_{n=0}^{c-1} w_n(c-c-n) = 0 \quad (i=2,3,...,c) \] (15b)

Eq. (15a) is derived by the expression \( \lim (z \to \infty) R(z) = 1 \). Then, the average queue length of customers in queue 1, \( E[L_1] \) is obtained as follows.

\[ E[L_1] = R(1) = \frac{\dot{A}(1)D(1) - A(1)\dot{D}(1)}{2(\dot{D}(1))^2} \] (16)

where,

\[ \dot{A}(1) = q \sum_{n=0}^{c-1} w_n(c-n) \]

\[ \dot{A}(1) = q \sum_{n=0}^{c-1} w_n(c(c-1)-n(n-1)) \]

\[ \dot{B}(1) = c-1 \sum_{n=0}^{c-1} \dot{w}_n(c-n) \]

\[ \dot{B}(1) = c-1 \sum_{n=0}^{c-1} \dot{w}_n(c-n) \]

3.2. Approximation method for \( q \)

Introducing the probability, \( q \) that customers in queue 1 can be served, the approximate probability generating function \( R(z) \) for the queue length of customers in queue 1 can be obtained in Section 3.1. Whether customers in queue 1 can be served depends on not only the state of queue 1 but also the states of other queues. So it is very difficult to calculate \( q \) exactly. Then, we propose the following approximation equation for \( q \) using \( c, b_k \) (the mean group size of customers in queue 1), \( \lambda_1 \) and \( \mu \) (the mean service rate).

\[ q = (\mu - \sum_{i=2}^{c} \frac{\lambda_i b_i}{\mu})/\mu = 1 - \sum_{i=2}^{c} \frac{\lambda_i b_i}{\mu} \] (18)

This equation is derived under the assumption that a server drops in queue 1, \( \mu \) times during \( c \) times service occasions and can collect customers in queue 1.

4. NUMERICAL EXAMPLES AND EVALUATIONS

We now evaluate the approximation method by comparing them with simulations.

Table 1 shows the validity of our simulation experiment, where simulation results are compared with exact theoretical values the case that the number of attributes is equal to one (i.e., HBQS case), and the length of service occasions is equal to one (i.e., HBQS case), and the time intervals of service ocassion epochs are constant.

The following numerical results are derived in the case that the sizes, \( B_i \) of customer groups at their arrival time are according to a geometric distribution with the mean equal to \( b_i \) \((=1/p)\), i.e.,

\[ B_i = Pr[B_i=k] = p(1-p)(k-1) \quad (k=1,2,...) \]

At first, we show the numerical results in the case where the time intervals of service occasions are according to an exponential distribution with parameter and the number of attributes is equal to two.

Fig. 4 shows the relationship between the average queue length of customers in queue 1, \( E[L_i] \) \((i=1,2)\) and the Poisson arrival rate of groups of customers whose attribute is \( i \), \( \lambda_i \) in the case where \( \lambda_1=0.1, b_1=b_2=5.0, \mu = 1.0 \) and \( c = 5 \). Lines and symbols present analytically obtained results and simulation results respectively. From this figure, we find that the theoretical results are well verified by the simulation results. Further it is found that \( E[L_1] \) is affected by the change of \( \lambda_1 \) but \( E[L_2] \) is not so much. Namely, in such a case, the mean queue length of one queue is not much

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Comparison between simulation and exact.</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \lambda )</td>
<td>0.1</td>
</tr>
<tr>
<td>Simulation</td>
<td>0.45</td>
</tr>
<tr>
<td>Exact</td>
<td>0.45</td>
</tr>
</tbody>
</table>

\( \mu = 1.0, b = 3.0, c = 3 \).
influenced by the arrival rate of group-customers in other queue.

Fig. 5 shows the relation between $E[L_1]$ and $c$ in the case where $\mu = 1.0$, $\lambda_2 = 0.1$ and $b_1 = b_2 = 5.0$. As illustrated in this figure, it is recognized that if $c$ is sufficiently larger than $b_1 = b_2$, $E[L_1]$ changes little. That is because all waiting customers in the same queue are considered to be served at once.

Next, we consider the case where the service occasion epochs have constant time intervals and the number of attributes is three.

In Fig. 6 the relationship between $E[L_1]$ and $\lambda_1$ are shown with $\lambda_2 = \lambda_3 = \lambda_4 = 5$ and $b_1 = b_2 = 5.0$ under various value of $b_1 = b_2 = b_3$. It is confirmed that slight increase of the bulk size of arriving customers causes the large queue length in the case of heavy traffic. Furthermore, even if the mean arrival number per time (i.e. $\lambda_1 b_1$) is the same, the larger the bulk size becomes, the more the queue length increases.

Fig. 7 shows the mean queue length of each queue as a factor of $b_1$. Similarly to Fig. 4, we can say that the mean queue length of one queue is not so much influenced by the mean customer-group size in other queue.

Fig. 8 shows the mean queue length for various $\lambda_1$ in the case where $\lambda_1 b_1$ is fixed to 0.4, $b_2 = b_3 = 4.0$, $c = 5$, $\mu = 1.0$, $b_1 = b_2 = 3.0$. The fitness of our approximation is not good for the case when the average bulk size of another attribute is too large.

In general, our approximation results are well verified by the simulation results shown in these figures but has tendency to underestimate the simulation ones, which is caused by the approximation of the probability, $q_k$, that customers in queue $i$, if any, can be served next derived. In eq. (18), we assumed $\lambda_1 b_1 / c$ as the mean frequency of service per time for queue $i$, but this value is smaller than exact one because a server does not always serve $c$ customers. Therefore, the value, $q$, of our method is greater than the real one, then mean queue length is underestimated.

Many examples show our approach provides approximation results with rather good accuracy, and, in general, has the similar tendency with these examples.

5. CONCLUSION

In this paper, we consider a transportation type bulk service queueing system with the composite service discipline as a fundamental research on future communication networks. In our model, waiting customers which have the same attribute can be served at the same time. Then, the number of the waiting customers with the same attribute is the key factor for performance of the systems. In our approach, each class of customers with the same attribute is assumed to form their own queue. An approximation method for obtaining queue length of each queue just before the service occasion epoch, where probability generating function approach and embedded Markov chain method are utilized. Many examples show that our proposed method provides high approximation accuracy evaluating by the comparison with the results of simulations. With the proposed method, it has become possible to evaluate the characteristic quantities of the system considered.

Our proposed method can be applied to analyze some of stochastic behaviors in communication system.

References


Fig. 4 Average queue length characteristics for various $\lambda_1$. 

\[ \mu = 1.0 \]
\[ c = 5 \]
\[ \lambda_2 = 0.1 \]
\[ b_1 = b_2 = 3.0 \]

$\Delta$: Simulation for queue 1
$\bigcirc$: Simulation for queue 2
Fig. 5 Average queue length characteristics for various capacity of a server.

\[ \mu = 1.0 \]
\[ \lambda_1 = \lambda_2 = 0.1 \]
\[ b_1 = b_2 = 5.0 \]
\[ \Delta : \text{Simulation} \]

Fig. 7 Average queue length characteristics for various \( b_1 \).

\[ \mu = 1.0 \]
\[ c = 5 \]
\[ \circ : \text{Simulation for } b_1 = 5.0 \ (i = 1, 2, 3) \]

Fig. 6 Average queue length characteristics for various \( \lambda \) and \( b \), where \( \lambda_1 = \lambda_2 = \lambda_3 \), and \( b_1 = b_2 = b_3 \).

\[ \lambda_1 b_1 = 0.6 \]
\[ \lambda_1 b_1 = 0.45 \]
\[ \lambda_1 b_1 = 0.40 \]

Fig. 8 Average queue length characteristics for various \( \lambda_1 \), where \( b_1 \) is inversely proportional to \( \lambda_1 \).