Mathematical model and numerical simulation of dual-channel multiphase bulk arrival queuing system with heterogeneous service time

Atina Ahdika
Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Islam Indonesia, Jalan Kaliurang Km 14.5 Sleman, Yogyakarta 55584, Indonesia
E-mail: atina.a@uii.ac.id

Abstract. There are many queue models have been developed by many researchers. Many previous studies have developed queue model for large system. This paper constructs a queue model for small system which consists of dual server with multiphase in one system. Each server has different service time and the customer comes in batch (bulk arrival). The model has been derived based on the concept of birth-death process and solved using global balance condition to obtain performance measurements of the system. Numerical simulation is carried out to get a better understanding about the correlation between each variable involved in the system. The result shows that the more dense the movement of the customers in the system, the more people and the longer time spent by a customer in the system.

1. Introduction
Queue model is a mathematical model of waiting lines which is constructed so that the time length spent in the line can be predicted. Many events in the real life can be modelled using queue model such as problems related to industrial manufactures, network systems, traffic systems, biological systems, etc. Researchers have developed many models which gave an optimal solution to minimize the number of queues and the length time of the queue in a system by adding the server or by accelerating the service rate in each server. Li and Stanford [1] developed a multi-class multi-server queuing model with heterogeneous servers under the accumulating priority queuing discipline. Sikdar and Gupta [2] developed a finite-buffer batch arrival and batch service queue with single and multiple vacations, they also derived various performance measures of the model. Ekpenyong and Udoh [3] extended and improved on the performance measure of the single-server single-queue system with multiple phases under the conditions of FCFS, infinite population source, Poisson arrival, and Erlang service time. Ghimire et al. [4] constructed a single server bulk queueing model with fixed batch size \( b \) and customers arrive to the system with Poisson rate \( \lambda \) and served exponentially with the rate \( \mu \). Maurya [5] studied a mathematical model of two-state batch arrivals and batch service Markovian queue with server vacation model under transient state. Sultan et al. [6] extended a Monte Carlo study to perform analysis of a multi-server bulk arrival queuing system with two-modes server breakdown. Sultan [7] introduced an easily applicable algorithm to solve problems involving bulk-arrival queues with a breakdown of one of the heterogeneous servers in case of steady state.
Most of the models discussed before are the model for a large queue system. While for small queue system is still rarely discussed. Whereas this small queue system can be helpful for small office or home-based industries. The models for large queue system have fairly high complexity and many conditions that must be met by the system. Departing from these matter, to simplify the model so that the model can be implemented easier for small queue system, then comes the idea to build a mathematical model and numerical simulation to the small queue system which is a combination of many cases. The model is a dual-channel multiphase bulk arrival queuing system with heterogeneous service time and infinite customer capacity. The arrival of the batch in this model is assumed as Poisson distributed and the service rate is exponentially distributed.

2. System construction and mathematical model

Queuing model built in this study is a model with two-channel services, $k$ phases of service that must be passed by every arrival, and the customer comes in the form of groups or large quantities. The construction of the model is described in Figure 1.

![Figure 1. Construction of the Model](image1)

The system shows that each arrival comes in batch, however customer leaves the system individually. The queue model can be met in home-based industry, service process in district office, new student registration system, or in the airport with two arrival gates where the arrival can comes in a group. The service rate in each channel is assumed to be different, so that each individual leaves the system in accordance with the minimum service rate. However, if only left one arrival, then the arrival will go to the channel with the fastest service rate. Suppose random variables represent the service rate $X_i \sim EXP(\mu_i)$, $i = 1, 2$. To determine which channel has faster service rate, we defined a new random variable $Y = min(X_1, X_2) \sim EXP(\mu_1 + \mu_2)$. The construction is developed from the smaller size of batch (bulk), notated by $b$. Figure 2 shows the process diagram of the constructed model.

![Figure 2. Process Diagram of the Constructed Model](image2)
\( \lambda \) is the arrival rate, \( k \) is the number of phases, \( 2k \sum \mu_i \) is total service rate of all phases, and \( k \cdot min(\mu_1, \mu_2) \) is total service rate of the faster channel if there is only one customer left. Probability function of the model is obtained by solving the global balance equation as follows:

\[
\begin{align*}
    k \cdot min(\mu_1, \mu_2)P_1 - \lambda P_0 &= 0 \\
    2k \sum_{i=1}^{2} \mu_i P_2 - [\lambda + k \cdot min(\mu_1, \mu_2)] P_1 &= 0 \\
    2k \sum_{i=1}^{2} \mu_i P_{n+1} - \left[ \lambda + 2k \sum_{i=1}^{2} \mu_i \right] P_n &= 0, \quad n = 2, 3, \ldots, b - 1 \\
    \lambda P_{n-b} + 2k \sum_{i=1}^{2} \mu_i P_{n+1} - \left[ \lambda + 2k \sum_{i=1}^{2} \mu_i \right] P_n &= 0, \quad n \geq b
\end{align*}
\]

Eq. (1) to (4) are used to obtain the probability generating function (pgf) of \( N \) (the number of customers in the system) that is \( G(z) = \sum_{n=0}^{\infty} z^n P_n \), and we obtain

\[
G(z) = \frac{2k \sum_{i=1}^{2} \mu_i (zP_1 + P_0)}{2k \sum_{i=1}^{2} \mu_i - \lambda z \sum_{j=0}^{b-1} z^j}
\]

Eq. (5) is an implicit function in form of \( P_0 \) and \( P_1 \). The equation can be changed into explicit form by using the correlation between \( P_0 \) and \( P_1 \) from eq. (1)

\[
P_1 = \frac{\lambda}{k \cdot min(\mu_1, \mu_2)} P_0
\]

The value of \( P_0 \) is obtained by substituting eq. (6) into eq. (5) at \( z = 1 \), where \( G(1) = \sum_{n=0}^{\infty} P_n = 1 \) and we obtain

\[
P_0 = 1 - \frac{\rho}{2} - P_1 = \frac{k \cdot min(\mu_1, \mu_2)[2 - \rho]}{2[\lambda + k \cdot min(\mu_1, \mu_2)]}
\]

where \( \rho = \lambda b/k \sum_{i=1}^{2} \mu_i \) represents the traffic intensity of the queue model for the case of bulk arrival and heterogeneous service rate. Then substitute eq. (7) into eq. (5) and obtained the pgf

\[
G(z) = \frac{k \sum_{i=1}^{2} \mu_i [2 - \rho]}{[\lambda + k \cdot min(\mu_1, \mu_2)]} \frac{[z \lambda + k \cdot min(\mu_1, \mu_2)]}{2k \sum_{i=1}^{2} \mu_i - \lambda \sum_{j=1}^{b-1} z^j}
\]

Probability of a certain amount of customers in the system can be calculated by differentiating \( G(z) \) \( n \) times at \( z = 0 \) and divided it by \( n! \), ie

\[
P(N = n) = \frac{1}{n!} \cdot \frac{d^n}{dz^n} G(z) \mid_{z=0}
\]
3. Performance measurements

3.1. Average Number of Customers in the System

The first performance measure is the average number of customers in the system, notated by \( L_S \). \( L_S \) is obtained by calculating the expectation of \( N \). However, we have probability generating function (pgf) instead of probability density function (pdf) of \( N \), therefore the expectation value is obtained by differentiating pgf in eq. (8) one time at \( z = 1 \).

\[
L_S = \sum_{n=0}^{\infty} n P_n = \frac{d}{dz} G(z) |_{z=1} = \frac{\rho k \sum_{i=1}^{2} \mu_i}{\rho k \sum_{i=1}^{2} \mu_i + kb \cdot \min(\mu_1, \mu_2)} + \frac{\rho (b + 1)}{2[2 - \rho]} \tag{10}
\]

3.2. Average Amount of Time a Customer Spent in the System

The average amount of time a customer spent in the system is notated by \( W_S \). The smaller the value of \( W_S \), the more efficient the existing queue system. The value of \( W_S \) is obtained by dividing average number of customer in the system with the total arrival rate.

\[
W_S = \frac{L_S}{\lambda b} = \frac{1}{\rho k \sum_{i=1}^{2} \mu_i + kb \cdot \min(\mu_1, \mu_2)} + \frac{(b + 1)}{2k \sum_{i=1}^{2} \mu_i[2 - \rho]} \tag{11}
\]

3.3. Average Amount of Time a Customer Spent Waiting in the Queue

The average amount of time a customer spent waiting in the queue is notated by \( W_Q \). \( W_Q \) is calculated only when a customer waiting in the queue not including when they are being served. The value of \( W_Q \) is obtained by subtracting the value of \( W_S \) with the expected value of the service time \( E(S) = 1/k \sum_{i=1}^{2} \mu_i \), thus

\[
W_Q = W_S - E(S) = \frac{1}{\rho k \sum_{i=1}^{2} \mu_i + kb \cdot \min(\mu_1, \mu_2)} + \frac{(b - 3 + 2\rho)}{2k \sum_{i=1}^{2} \mu_i[2 - \rho]} \tag{12}
\]

3.4. Average Number of Customers Waiting in the Queue

The last measure is the average number of customer waiting in the queue, notated by \( L_Q \). Similar with \( W_Q \), \( L_Q \) is only calculated based on the number of customers waiting in the queue. The value of \( L_Q \) is obtained by multiplying the average amount of time a customer waiting in the queue with the total arrival rate.

\[
L_Q = \lambda b W_Q = \frac{\rho k \sum_{i=1}^{2} \mu_i}{\rho k \sum_{i=1}^{2} \mu_i + kb \cdot \min(\mu_1, \mu_2)} + \frac{\rho (b - 3 + 2\rho)}{2[2 - \rho]} \tag{13}
\]

4. Numerical simulation

In this section, numerical simulation is conducted to explain correlation among the performance measurements so that it can be used as a consideration in determining the efficient system. The system called an efficient system if the value of traffic intensity is less than one, \( \rho < 1 \), i.e. \( \rho = \lambda b/k \sum_{i=1}^{2} \mu_i < 1 \) so that \( \lambda b < k \sum_{i=1}^{2} \mu_i \).
In other words, when the arrival rate is less than the service rate, then there will be no solid queue in the system. The relation between $\rho$ and the performance measurements for different size of arrival $b$ is discussed in this section. The simulation was conducted by setting the value of $\rho$ from 0.1 to 1 and generated the value of $n$, $\mu_1$, $\mu_2$, and $k$ for various value of $b$. The value of $\lambda$ is represented by the value of $\rho$. Table 1 shows the variables generated in the simulation.

Table 1. Generated Variables of the Simulation

| Simulation | $\rho$          | $b$          | $n$ | $k$ | $\mu_1$ | $\mu_2$ |
|-----------|-----------------|--------------|-----|-----|---------|---------|
| 1         | 0.1, 0.2, ..., 0.9, 1 | 5, 10, 15, 20, 25 | 50  | 5   | 7       | 8       |
| 2         |                 | 15, 30, 45, 60, 75 | 200 | 10  | 32      | 28      |
| 3         |                 | 10, 20, 30, 40, 50 | 100 | 5   | 10      | 15      |
| 4         |                 | 25, 50, 75, 100, 125 | 300 | 10  | 30      | 25      |

The result of the simulations (are only shown for simulation 1 and 2 because the rest provide the same behaviour), the correlation between $\rho$, $b$, $L_S$, $L_Q$, $W_S$, and $W_Q$ are shown by Figure 3 and 4.

Figure 3 and 4 show that the larger the value of $\rho$, the larger the value of the performance measurements $L_S$, $L_Q$, $W_S$, and $W_Q$ for various value of $b$, $n$, $k$, $\mu_1$, and $\mu_2$. It is indicated that the more dense the movement of customers in the queue, the more people who are in the system or in the queue and also the longer the time spent by a customer in the system or in the queue. In addition, the larger the size of arrival $b$, the greater the values of $L_S$, $L_Q$, $W_S$, and $W_Q$. From the simulations, it can also be seen that Simulation 2 has the shortest time spent by a customer in the system and in the queue. It is happened because Simulation 2 has the fastest service rate among others. Based on the results of the numerical simulation, it is concluded that the model can be implemented for small queue system and the system can shortened the queue by accelerating the service rate of the servers.
5. Conclusion

The construction of a dual-channel multiphase bulk arrival queuing system with heterogeneous service time has been developed. It shows that the mathematical model in form of explicit pgf can be derived. Performance measurements of the model have also been obtained. The correlation between variables involved in the model with the performance measurements has been shown by the numerical simulation. Overall, it can be concluded that the traffic intensity $\rho$ is directly proportional with the performance measurements and the size of the arrival $b$. The model which has been constructed can be implemented to small queue system.

6. References

[1] Na Li and David A. Stanford, 2016, Multi-server Accumulating Priority Queues with Heterogeneous Servers, *European Journal of Operational Research* 252 866-878.

[2] K. Sikdar and U.C. Gupta, 2008, On the Batch Arrival Batch Service Queue with Finite Buffer Under Server’s Vacation: $M^X/G^Y/1/N$ Queue, *Computers and Mathematics with Application* 56 2861-2873.

[3] Emmanuel John Ekpenyong and Nse Sunday Udoh, 2011, Analysis of Multi-Server Single Queue System with Multiple Phases, *Pak.j.stat.oper.res. Vol VII No.2* 305-314.

[4] Sushil Ghimire, R.P. Ghimire, and Gyan Bahadur Thapa, 2014, Mathematical Models of $M^b/M/1$ Bulk Arrival Queueing System, *Journal of the Institute of Engineering Vol. 10 No. 1* 184-191.

[5] Vishwa Nath Maurya, 2014, Mathematical Modelling and Performance Analysis of Single Server Two-State Batch Arrivals and Batch Service Markovian Queue with Multiple Vacations, *American Journal of Modeling and Optimization Vol. 2 No. 2* 239-246.

[6] A.M. Sultan, N.A. Hassan, and N.M. Elhany, 2005, Computational Analysis of A Multi-Server Bulk Arrival with Two Modes Server Breakdown, *Mathematical and Computational Applications Vol. 10 No. 2* 249-259.

[7] Ahmed M.M. Sultan, Multi-Channel Bi-Level Heterogeneous Servers Bulk Arrival Queueing System with Erlangian Service Time, *Mathematical and Computational Applications Vo. 12 No. 2* 97-105 (2007).