Mathematical Theory and Modeling
ISSN 2224-5804 (Paper) ISSN 2225-0522 (Online)
Vol.11, No.5, 2021

On single server batch arrival queueing system with balking, three types of heterogeneous service and Bernoulli schedule server vacation

Samuel U. Enogwe[1*], Sidney I. Onyeagu[2], Happiness O. Obiora-Ilouno [3]

[1]Department of Statistics, Michael Okpara University of Agriculture, Umudike, Nigeria
[2,3] Department of Statistics, Nnamdi Azikiwe University, Awka, Nigeria
Corresponding Author: senogwe@yahoo.com[1*] / enogwe.samuel@mouau.edu.ng ; https://orcid.org/0000-0003-1100-7198

Abstract

This paper investigates a batch arrival queueing system in which customers arrives at the system in a Poisson stream following a compound Poisson process and the system has a single server providing three types of general heterogeneous services. At the beginning of each service, a customer is allowed to choose any one of the three services and as soon as a service of any type gets completed, the server may take a vacation or may continue staying in the system. The vacation time is assumed to follow a general (arbitrary) distribution and the server vacation is based on Bernoulli schedule under a single vacation policy. During the server vacation period, impatient customers are assumed to balk. This paper described the model as a bivariate Markov chain and employed the supplementary variable technique to find closed-form solutions of the steady state probability generating function of number of customers, the steady state probabilities of various states of the system, the average queue size, the average system size, and the average waiting time in the queue as well as the average waiting time in the system. Further, some interesting special cases of the model are also derived.
Keywords Batch Arrivals. Queueing System. Balking. Heterogeneous types of Service. Bernoulli schedule server vacation. Bivariate Markov Processes.

MSC2020-Mathematics Subject Classification 34B07, 60G05, 62E15

1 Introduction

On daily basis, customers (human beings or physical entities) that need certain kind of service go to a service centre to receive such service (Enogwe, 2021). Usually, the service facility contains mechanism(s) called server(s) or service channel(s), which perform the service on customers (Enogwe, 2020). If the number of customers exceed the number of service facilities or the service facilities do not work efficiently and take more time than prescribed to serve a customer, then a queue is formed (Sharma, 2013).

A queue is a line of customers that are waiting to be served in a particular service centre. Queues are seen in many places. For example, customers queue-up in the bank to make deposits, customers line-up in a canteen to buy food, vehicles wait in line at a petrol station for refuelling, students wait in line to be screened for admission, voters' queue-up to cast their voters, customers wait in line in a supermarket to pay for groceries, patients wait in line in a hospital to see a doctor, machines wait in workshops to be repaired, mechanics wait in workshop to receive tools, trucks wait in warehouses to be unloaded or offloaded, vehicles wait on road for traffic lights, airplanes wait to take-off or land, order waits to be processed, electronic messages wait to be delivered, a train wait at outer signal for green signal, among others (Sharma, 2013; Gupta and Hira, 2013; Murthy, 2007; Hillier and Lieberman, 2001).

Formation of queue is a common phenomenon whether it is visible (in case of human beings) or invisible (in case of inanimate object), but waiting in queue is one of the unpleasant and undesirable activities of life. Obviously, waiting is frustrating, demoralizing, agonizing, aggravating, annoying, time consuming and incredibly expensive. Great inefficiencies also occur because of other kinds of waiting than people standing in line. For example, machines waiting to be repaired may result in lost production. Vehicles that need to wait to be unloaded may delay subsequent shipments. Airplanes waiting to take off or land may disrupt later travel schedules. Delays in telecommunication transmissions due to saturated times may cause data glitches. Causing manufacturing jobs to wait to be performed may disrupt subsequent production. Delaying service jobs beyond their due dates may result in lost future business (Hillier and Lieberman, 2001).

In order to reduce the length of queue in a service centre, the service capacity has to be increased. The method of reducing congestion by the expansion of service capacity may result in an increase in idle time of the service station and may become uneconomical for the organization. In addition, providing too much service capacity to operate the system involves high costs. But not providing enough service capacity results in excessive waiting and all its unfortunate consequences.

Queueing theory is the mathematical study of "queues" or "waiting" lines. It uses queueing models to represent the various types of queueing systems that arise in practice. Formulae for each queueing model indicate how the corresponding queueing system should perform, including the average amount of waiting that will occur under a variety of circumstances. Further, queueing theory helps managers to reduce the waiting time of customers and suggests to the organization optimal number of service facilities to install, so that customer will be happy and the organization can run the business economically. The theory tries to strike a balance between the costs associated with waiting and costs of preventing waiting and helps analysts to determine the optimal number of service facilities required and optimal arrival rate of the customers of the system (Murthy, 2007).

A classical queueing system is composed of customers or units needing some kind of service who arrive at a service facility where such service is provided, join a queue if service is not immediately available, and eventually leave after receiving service (Medhi, 2003).

Two commonly used queueing models are Markovian and non-Markovian queueing models (Medhi, 2003). The queueing models in which both the interarrival time and the service time distributions are exponential are birth-death Markovian whereas models in which the distributions of either the interarrival time or the service time distributions or both are Erlangian are non-birth-death but nevertheless can be treated as Markovian. Markovian queues are well developed and excellent accounts of Markovian queueing theory can be found in Medhi (2003) and Sharma (2013) among many other standard books. Markovian queueing models are widely used in many fields because these models can be analysed with considerable ease and they provide fairly good approximations for the interarrival time and the service time distributions. Queueing models having interarrival times and service times which are not exponential or Erlangian are called non-Markovian queueing models. A particular class of non-Markovian queueing models which has been found useful in many fields is the batch arrival single-server queueing model.

A batch arrival single server queueing system is one in which customers arrive for service in groups or in batches and are served individually by one server. Some examples of batch arrivals include families that go to a restaurant for lunch in a particular period of the day, convoy of a political officer that came to a filling station for refuelling and sets of triplets brought to a hospital for medical treatment. Other examples are batch of raw materials supplied to an industry for manufacturing, a group of imported items to be unloaded at a warehouse and so on. Batch arrival queueing models have found applications in telecommunication systems, banks, communication systems and large-scale manufacturing industries, to mention a few and have been extensively discussed in queueing theory literature. One could check Bailey (1954), Cox (1955), Keilson and Kooharian (1960), Borthakur and Medhi (1974), Ross (1980), Cohan (1982), Whitt (1983) for further details.

Another class of queueing system commonly used in practice is the vacation queueing system. A vacation queueing system is one in which a server may become unavailable for random period of time from a primary service centre. The time away from the primary service centre is called a vacation, and customers who arrive while the server is on vacation will have to wait until the server returns from vacation. A vacation can be the result of many factors. In some cases, the vacation can be the result of server breakdown, which means that the system must be repaired and brought back to service. It can also be a deliberate action taken to utilize the server in a secondary service centre when there are no customers present at the primary service centre. Thus, server vacations are useful for those systems in which the server's idle time is utilized for other purposes, and this makes the queueing model to be applicable to a variety of real-world stochastic service systems (Ibe, 2015). Applications that can be modelled by the vacation queueing systems include machine breakdowns and maintenance in communication and computer systems, production and manufacturing systems, airline scheduling, inventory systems among others.

In a bid to developing mathematical queueing models, it is often assumed that the server is always available for providing service to customers. However, in practice, this assumption is not always true because the server may be unavailable for some time due to the fact that it is on vacation. During vacation time, the server may be performing supplementary job or it may be undergoing maintenance or it may simply take a break. Allowing the server to take vacations makes the queueing model more realistic and flexible in studying the real-life queueing problems. Vacation queueing models find applications in places like call centres with multi-task employees, customized manufacturing, telecommunication, and computer networks problems. Apart from the work of Levy and Yechiali (1975), which marks the beginning of vacation queueing models, several studies have been subsequently undertaken within the context of vacation queueing models by researchers like (Baba (1986); Borthakur and Choudhury (1997); Chao and Zhao (1998); Choi and Park (1990); Choudhury (2000, 2002); Choudhury and Madhuchanda (2005); Cramer (1989); Doshi (1986); Fuhrman (1981); Ibe (2015); Keilson and Servi (1987); Lee *et al.* (1995); Li and Zhu (1996); Madan and Choudhury (2005); Rosenberg and Yechiali (1993); Shanthikumar (1988); Tagaki (1991); Zhang and Vickson (1993)), among others.

Another essential, but limiting, assumption of classical batch arrival queueing models is that there is no balking. By "balking," we mean a situation whereby customers refused to join the queue either by seeing the number of customers already in queue (i.e., long length of the queue) or by estimating the duration of waiting time for a service to get completed. However, balking is, in fact, a usual practice in queueing systems because some customers upon seeing the length of the queue decide not to join the waiting line. For example, a customer that wants to go by train to his destination goes to railway station and after seeing the long queue in front of the ticket counter, may not like to join the queue and seek other type of transport to reach his/her destination. We see applications of queues with balking in emergency units in hospitals dealing with critical patients, communication systems, banks, production and inventory systems and many more. Queueing models with balking have been discussed by (Altman and Yechiali (2006); Barrer (1957); Baruah *et al.* (2013); Doshi (1991); Hagighi *et al.* (1986); Haight (1957); Kumar and Sharma (2012); Madan (2002)), and many others.

Moreover, studies by (Choudhury (2002); Choudhury and Mandhuchanda (2005) among several others) have a common assumption that the system has a single server who provides only one kind of service to the arriving customers. In order to obtain a more realistic model, Madan et al. (2005) proposed and studied a batch arrival queue with a single server providing two kinds of general heterogeneous service. In a related study, Anabosi and Madan (2003) introduced a single server queue with two types of service, optional server vacations based on Bernoulli schedule and single vacation policy, where vacation time was assumed to be exponentially distributed. Madan *et al.* (2005) extended the work of Anabosi and Madan (2003) to a case where both service and vacation times followed general and arbitrary distributions. Ebenesar Anna Bagyam and Udaya Chandrika (2011) developed a single server queueing model with impatient customers whereby the server provides two types of service and each arriving customer has the option of choosing either type of service. Baruah *et al.* (2012) investigated a batch arrival vacation queue with balking and re-service as well as two types of heterogeneous service. Ebenesar Anna Bagyam and Udaya Chandrika (2013) analysed a single server batch arrival retrial queueing system with optional extended server vacation, where server provides two stages of heterogeneous service in succession and each phase has two types of service that could be selected by the customers. Maragathasundari and Srinivasan (2012) carried out a transient analysis of single server batch arrival queue with Bernoulli feedback and three types of service. Another study by Maragathasundari et al. (2013) dealt with a batch arrival non-Markovian queueing model with three types of service and the server was assumed to follow multiple vacation policy. Recently, Mahanta and Choudhury (2018) introduced a non-Markovian batch arrival multiple vacation queue with two types of heterogeneous services with Bernoulli feedback.

In the literature relating to queueing theory, we do not find any single server model that combines batch arrivals, three types of heterogeneous service, balking and Bernoulli schedule server vacation. Motivated by the wide applicability of such a model, we propose a model which combines all these characteristics. This model has potential application in a flexible manufacturing system, where the parts to be processed arrives to a workstation in batches of random size instead of single units. A workstation may be in charge of three types of services at once. So, the vacation in our model may correspond to the time duration it is working on other secondary jobs such as maintenance work or simply taking rest. During the vacation period, several other customers may arrive and create a long queue. If any further customer arrives and refuses to enter the system on seeing the long length of the queue which resulted because the server is on vacation, then balking is said to have occurred. For this reason, our model has incorporated balking in its assumptions. Notably, server vacations, the types of services rendered by the server and the balking behaviour of customers have significant effects on system performance. The rest of the paper is organized as follows. Sect. 2 deals with mathematical description of the proposed queueing model. Definitions, notations and equations governing the proposed queueing model are given in Sect. 3. In Sect. 4 the results of the proposed queueing model are given. Some special cases of the proposed model are given in Sect. 5. The conclusion of this work is given in Sect. 6.

2 Mathematical description of the proposed queueing model

Queueing systems are described according to Kendall's notation for naming queueing systems. To this end, the proposed queueing model is denoted by $M^{[Y]}/\left(G_1G_2G_3\right)^T/1/G\left(BS\right)/V_s/B_{lk}$, where $\left(G_1G_2G_3\right)^T$ stands for three types of general heterogeneous service (one of which has to be chosen by each customer), $G\left(BS\right)$ denotes general service time under Bernoulli schedules and V_s denotes single vacations and B_{lk} denotes "balking" customer behaviour.

The following assumptions briefly describe the mathematical model being considered in this paper:

(i) Customers arrive at the system in batches of variable size according to a compound Poisson process with arrival rate, λ. Let the arrival batch size Y be a random variable with probability mass function $P\left(Y=i\right)=c_i$, $i=1,2,3,...$, then $\lambda c_i\Delta y\left(i=1,2,3,...\right)$ denotes the first order probability that a batch of i customers arrives at the system during a short interval of time $\left(y,\ y+\Delta y\right]$, where $0\le c_i\le 1,\ \sum_{i=1}^{\infty}c_i=1$.

(ii) There is a single server providing three types of general heterogeneous service to customers one by one on a first come, first served (FCFS) basis. Before a service starts, each customer can choose type 1 service with probability ξ_1 or type 2 service with probability ξ_2 or type 3 service with probability ξ_3, where $\xi_1+\xi_2+\xi_3=1$. The service times of the three types of services are assumed to follow different general (arbitrary) distributions. Let $G_j\left(y\right);\ j=1,2,3$ and $g_j\left(y\right);\ j=1,2,3$ be the distribution function and density function of the three types of heterogeneous services respectively. The conditional probability density of completion of the jth type of service during the interval $(y, y+dy]$, given that the elapsed service time is y, is given by $\mu_j(y)dy$, so that

$$\mu_j(y) = \frac{g_j(y)}{1-G_j(y)} \quad , j = 1,2,3 \tag{1}$$

$$\Rightarrow g_j(y) = \mu_j(y) e^{-\int_0^y \mu_j(s)ds} \quad , \; j = 1,2,3 \tag{2}$$

(iii) Once the service of a customer is completed, the server may go on a vacation of random length of time with probability p, where $0 \le p \le 1$ or it may stay in the system to serve the next customer, if any, with probability $(1-p)$, otherwise, it may remain idle in the system if there is no customer requiring service. The vacation times are assumed to follow a general (arbitrary) distribution with distribution function $W(y)$ and probability density function $w(y)$, respectively. Let $\phi(y)dy$ be the conditional probability of density of completion of a vacation period during the interval $(y, y+\Delta y]$, given that the elapsed vacation time is y, so that

$$\phi(y) = \frac{w(y)}{1-W(y)} \tag{3}$$

$$\Rightarrow w(y) = w(y) e^{-\int_0^y \phi_j(v)dv} \tag{4}$$

(iv) An arriving batch of customers balks during the period when the server is busy (available on the system) with probability $(1-\omega_1),(0 \le \omega_1 \le 1)$, so that ω_1 denotes the probability that an arriving batch of customers joins the system at the time when the server is busy. Also, it is assumed that an arriving batch of customers balks during the period when the server is on vacation with probability $(1-\omega_2),(0 \le \omega_2 \le 1)$, so that ω_2 denotes the probability that an arriving batch of customers of joins the system during the server vacation period.

(v) The interarrival times of customers, the service times of each type of service, the vacation time of the server as well as all stochastic processes involved in the queuing system are independent of each other.

3 Definitions, notations and equations governing the proposed queueing model

3.1 Definitions and Notations

Let $N_q(t)$ denote the queue size (excluding one in service) at time t, $V^0(t)$ be the elapsed vacation time of the server, and $G_j^0(t)$ be the elapsed service time of the customer for the jth type of service at time t, with $j = 1,2,3$, denoting First Type Service (FTS), Second Type

Service (STS), and Third Type Service (TTS), respectively. Now, we introduce the variable $Y(t)$ as follows:

$$Y(t)=\begin{cases}0 \text{ if the server is idle at time } t\\ 1 \text{ if the server is busy with first type service at time } t\\ 2 \text{ if the server is busy with second type service at time } t\\ 3 \text{ if the server is busy with third type service at time } t\\ 4 \text{ if the server is on vacation at time } t\end{cases} \tag{5}$$

Next, we introduce the supplementary variable as

$$X(t)=\begin{cases}0 & \text{if } Y(t)=0\\ G_1^0(t) & \text{if } Y(t)=1\\ G_2^0(t) & \text{if } Y(t)=2\\ G_3^0(t) & \text{if } Y(t)=3\\ V^0(t) & \text{if } Y(t)=4\end{cases} \tag{6}$$

Notice from (5) and (6) that we have treated the elapsed service time of customers for the three types of service and elapsed vacation time as supplementary variables. As stated in Kashyap and Chaudhry (1988), the supplementary variables are introduced to obtain a bivariate Markov process $\{N_q(t), X(t)\}$, which enables us to write the governing equations of the queueing model.

Let us assume that the system is in steady state condition. By steady state condition, we mean the normal condition that a queueing system is in after operating for some time with a fixed utilization factor less than one (Sharma, 2013). We then let $P_{n,j}(y)$ denote the steady state probability that the server is active providing type $j\,(j=1,2,3)$ service and there are $n\,(n\geq 1)$ customers in the queue excluding the one receiving type $j\,(j=1,2,3)$ service and the elapsed serviced time of this customer is y. Accordingly, $P_{n,j}=\int_0^\infty P_{n,j}(y)\,dy$ denotes the corresponding steady state probability that there are $n\,(n\geq 1)$ customers in the queue excluding the one receiving type $j\,(j=1,2,3)$ service irrespective of the elapsed service time y of this customer; $V_n(y)$ denotes steady state probability that the server is on vacation with elapsed vacation time y, and there are $n\,(n\geq 0)$ customers in the queue. Accordingly, $V_n=\int_0^\infty V_n(y)\,dy$ denotes the corresponding steady state probability that there are $n\,(n\geq 0)$ customers in the queue and the server is on vacation irrespective of the elapsed vacation time y of the server; and Q denotes steady state probability that there are no customers in the system and the server is idle but available in the system.

Next, we define the probability generating functions (PGFs) used in this paper as follows:

$$\left.\begin{array}{l} P_j(x,z)=\sum_{n=1}^{\infty} z^n P_{n,j}(y)\ ; P_j(z)=\sum_{n=1}^{\infty} z^n P_{n,j}\ ,\ j=1,2,3 \\ V(x,z)=\sum_{n=1}^{\infty} z^n V_n(y)\ ; V(z)=\sum_{n=1}^{\infty} z^n V_n \\ C(z)=\sum_{i=1}^{\infty} z^i c_i \end{array}\right\} \quad (7)$$

which are convergent inside the circle given by $|z| \le 1$. Also, to be defined in this paper is the Laplace-Stieltjes transform of a function $F(t)$:

$$F^*(s)=\int_0^{\infty} e^{-st} dF(t) \quad (8)$$

According to Medhi (2003), the probability generating functions and the Laplace-Stieltjes transforms are useful in finding solution to the set of differential difference equations governing the queuing system. In particular, they are used for finding closed form expressions of the probabilities of the states of the server.

3.2 Governing equations of the proposed queueing model

To derive the governing equations based on the assumptions stated in Section 2, we recognize that the following three events (possibilities) may occur in the system during a short interval of time $(y, y+\Delta y]$:

(i) There are n customers in the queue excluding one customer in type $j\ (j=1,2,3)$ service and elapsed service time is y, no arrivals and no service completion during the interval $(y, y+\Delta y]$. The joint probability for this event is given by $(1-\lambda\Delta y)(1-\mu_j(y)\Delta y)P_{n,j}(y)$.

(ii) There are n customers in the queue excluding one customer in service since the elapsed service time is y and an arriving batch of customers balks with probability $(1-\omega_1)$. The joint probability for this event is given by $(1-\omega_1)\lambda\Delta y P_{n,j}(y)$.

(iii) There are $(n-i)$ customers in the queue excluding the one in type $j\ (j=1,2,3)$ service since the elapsed service time is y, a batch of size i customers arrive and join the queue with probability ω_1. The joint probability for this event is given by $\omega_1\lambda\Delta y\sum_{i=1}^{n} c_i P_{n-i,j}(y)$.

Based on the three events stated above, we first find an expression for $P_{n,j}(y+\Delta y)$, known as the probability of n customers in the queue excluding one customer receiving type $j\ (j=1,2,3)$ service and the elapsed service time of the customer is $(y, y+\Delta y]$. To find the expression for $P_{n,j}(y+\Delta y)$, we sum the probabilities of the three mutually exclusive events enumerated in (i), (ii) and (iii) above. Consequently,

$$P_{n,j}(y+\Delta y) = P_{n,j}(y)(1-\lambda\Delta y)\left(1-\mu_j(y)\Delta y\right) + P_{n,j}(y)(1-\omega_1)(\lambda\Delta y) + \sum_{i=1}^{n-1} c_i P_{n-i}(x)(\omega_1\lambda\Delta y) \quad ; j=1,2,3; n\geq 1 \tag{9}$$

For $n=0$, (9) reduces to

$$P_{0,j}(y+\Delta y) = P_{0,j}(y)[1-\lambda\Delta y]\left[1-\mu_j(y)\Delta y\right] + P_{0,j}(y)[1-\omega_1][1-\lambda\Delta y]\mu_j(y),\ j=1,2,3 \tag{10}$$

Following the logic for obtaining (9) and (10), we get

$$V_n(y+\Delta y) = V_n(y)[1-\lambda\Delta y]\left[1-\phi(y)\right] + V_n(y)[1-\omega_2][\lambda\Delta y] + \omega_2\lambda\Delta y\sum_{i=1}^{n-1} c_i V_{n-i}(y) \quad , n\geq 1 \tag{11}$$

$$V_0(y+\Delta y) = V_0(y)[1-\lambda\Delta y]\left[1-\phi(y)\right] + V_n(y)[1-\omega_2][\lambda\Delta y] + \omega_2\lambda\Delta y\sum_{i=1}^{n-1} c_i V_{n-i}(y) \tag{12}$$

Similarly, we follow the logic used for obtaining (9) to get

$$Q = Q(1-\lambda\Delta y) + Q(1-\omega_2)(\lambda\Delta y) + (1-p)\sum_{j=1}^{3}\int_0^\infty P_{0,j}(y)\mu_j(y)\,dy + \int_0^\infty V_0(y)\phi(y)\,dy \tag{13}$$

By manipulating (9)-(13) as $\Delta y \to 0$, we obtain the steady state equations governing the proposed queuing model as follows:

$$\frac{d}{dy}P_{n,1}(y) + \left(\lambda+\mu_1(y)\right)P_{n,1}(y) = (1-\omega_1)\lambda P_{n,1}(y) + \omega_1\lambda\sum_{i=1}^{n} c_i P_{n-i}(y), n\geq 1 \tag{14}$$

$$\frac{d}{dy}P_{0,1}(y) + \left(\lambda+\mu_1(y)\right)P_{0,1}(y) = (1-\omega_1)\lambda P_{0,1}(y) \tag{15}$$

$$\frac{d}{dy}P_{n,2}(y) + \left(\lambda+\mu_2(y)\right)P_{n,2}(y) = (1-\omega_1)\lambda P_{n,2}(y) + \omega_1\lambda\sum_{i=1}^{n} c_i P_{n-i}(y), n\geq 1 \tag{16}$$

$$\frac{d}{dy}P_{0,2}(y) + \left(\lambda+\mu_2(y)\right)P_{0,2}(y) = (1-\omega_1)\lambda P_{0,2}(y) \tag{17}$$

$$\frac{d}{dy}P_{n,3}(y) + \left(\lambda+\mu_3(y)\right)P_{n,3}(y) = (1-\omega_1)\lambda P_{n,3}(y) + \omega_1\lambda\sum_{i=1}^{n} c_i P_{n-i}(y), n\geq 1 \tag{18}$$

$$\frac{d}{dy}P_{0,3}(y) + \left(\lambda+\mu_3(y)\right)P_{0,3}(y) = (1-\omega_1)\lambda P_{0,3}(y) \tag{19}$$

$$\frac{d}{dy}V_n(y) + \left(\lambda+\phi(y)\right)V_n(y) = \lambda(1-\omega_2)V_n(y) + \omega_2\lambda\sum_{i=1}^{n} c_i P_{n-i}(y), n\geq 1 \tag{20}$$

$$\frac{d}{dy}V_0(y) + \left(\lambda+\phi(y)\right)V_0(y) = \lambda(1-\omega_2)V_n(y) + \omega_2\lambda\sum_{i=1}^{n} c_i V_{n-i}(y) \tag{21}$$

$$Q=Q(1-\lambda)+\lambda Q(1-\omega_2)Q+(1-p)\begin{bmatrix}\int_0^\infty P_{0,1}(y)\mu_1(y)dy+\int_0^\infty P_{0,2}(y)\mu_2(y)dy\\+\int_0^\infty P_{0,3}(y)\mu_3(y)dy\end{bmatrix}+\int_0^\infty V_0(y)\phi(y)dy \tag{22}$$

Notably, Equations (13)-(22) are to be solved subject to the following boundary conditions:

$$P_{n,1}(0)=(1-p)\xi_1\begin{bmatrix}\int_0^\infty P_{n+1,1}(y)\mu_1(y)dy+\int_0^\infty P_{n+1,2}(y)\mu_2(y)dy\\+\int_0^\infty P_{n+1,3}(y)\mu_3(y)dy\end{bmatrix}+\xi_1\int_0^\infty V_{n+1}\phi(y)dy+\lambda\omega_1\xi_1c_{n+1}Q\ ,n\geq 0 \tag{23}$$

$$P_{n,2}(0)=(1-p)\xi_2\begin{bmatrix}\int_0^\infty P_{n+1,1}(y)\mu_1(y)dy+\int_0^\infty P_{n+1,2}(y)\mu_2(y)dy\\+\int_0^\infty P_{n+1,3}(y)\mu_3(y)dy\end{bmatrix}+\xi_2\int_0^\infty V_{n+1}\phi(y)dy+\lambda\omega_1\xi_2c_{n+1}Q\ ,n\geq 0 \tag{24}$$

$$P_{n,3}(0)=(1-p)\xi_3\begin{bmatrix}\int_0^\infty P_{n+1,1}(y)\mu_1(y)dy+\int_0^\infty P_{n+1,2}(y)\mu_2(y)dy\\+\int_0^\infty P_{n+1,3}(y)\mu_3(y)dy\end{bmatrix}+\xi_3\int_0^\infty V_{n+1}\phi(y)dy+\lambda\omega_1\xi_3c_{n+1}Q\ ,n\geq 0 \tag{25}$$

$$V_n(0)=p\left[\int_0^\infty P_{n+1,1}(y)\mu_1(y)dy+\int_0^\infty P_{n+1,2}(y)\mu_2(y)dy+\int_0^\infty P_{n+1,3}(y)\mu_3(y)dy\right]\ ,n\geq 0 \tag{26}$$

and the normalizing condition

$$Q+\sum_{j=1}^{3}\sum_{n=0}^{\infty}\int_0^\infty P_{n,j}(y)dy+\sum_{n=0}^{\infty}\int_0^\infty V_n(y)dy=1 \tag{27}$$

4 Results of the proposed queueing model

4.1 Queue size distribution at random epoch for the proposed queueing model

Theorem 1 *Under the stability condition* $\rho<1$, *the queueing model* $M^{[Y]}/(G_1G_2G_3)^T/1/G(BS)/V_s/B_{lk}$, *has the following marginal probability generating functions for the server's state queue size:*

$$P_1(z)=\frac{\xi_1\left[1-G_1^*(m)\right]Q}{B(z)} \tag{28}$$

$$P_2(z)=\frac{\xi_2\left[1-G_2^*(m)\right]Q}{B(z)} \tag{29}$$

$$P_3(z)=\frac{\xi_3\left[1-G_3^*(m)\right]Q}{B(z)} \quad (30)$$

$$V(z)=\frac{p\frac{\omega_1}{\omega_2}\left[1-W^*(n)\right]\left[\xi_1G_1^*(m)+\xi_2G_2^*(m)+\xi_3G_3^*(m)\right]Q}{B(z)} \quad (31)$$

where

$$B(z)=z-\left[(1-p)+pW^*(n)\right]\left[\xi_1P_1(0,z)G_1^*(m)+\xi_2P_2(0,z)G_2^*(m)+\xi_3P_3(0,z)G_3^*(m)\right] \quad (32)$$

Proof In order to derive Theorem 1, we multiply (14) by z^n, take the sum of both sides over n from 1 to ∞, add the resulting equation to (15) and utilize (7) to obtain

$$\frac{d}{dy}P_1(y,z)+\left\{\omega_1(\lambda-\lambda C(z))+\mu_1(y)\right\}P_1(y,z)=0 \quad (33)$$

Multiplying (16) by z^n and summing over n from 1 to ∞, adding the result to (17) and making use of (7), we have

$$\frac{d}{dy}P_2(y,z)+\left\{\omega_1(\lambda-\lambda C(z))+\mu_2(y)\right\}P_2(y,z)=0 \quad (34)$$

Multiplying (18) by z^n and summing over n from 1 to ∞, adding the result to (19) and using (7), we have

$$\frac{d}{dy}P_3(y,z)+\left\{\omega_1(\lambda-\lambda C(z))+\mu_3(y)\right\}P_3(y,z)=0 \quad (35)$$

Multiplying (20) by z^n and summing over n from 1 to ∞, adding the result to (21) and utilizing (7), we have

$$\frac{d}{dy}V(y,z)+\left\{\omega_2(\lambda-\lambda C(z))+\phi(y)\right\}V(y,z)=0 \quad (36)$$

Notice that (33)-(36) constitutes first order differential equations and can be solved by integration using separation of variables technique. Hence, integrating (33)-(36) between the limits 0 to y, we obtain

$$P_1(y,z)=P_1(0,z)e^{-\omega_1(\lambda-\lambda C(z))y-\int_0^y\mu_1(t)dt} \quad (37)$$

$$P_2(y,z)=P_2(0,z)e^{-\omega_1(\lambda-\lambda C(z))y-\int_0^y\mu_2(t)dt} \quad (38)$$

$$P_3(y,z)=P_3(0,z)e^{-\omega_1(\lambda-\lambda C(z))y-\int_0^y\mu_3(t)dt} \quad (39)$$

$$V(y,z)=V(0,z)e^{-\omega_2(\lambda-\lambda C(z))y-\int_0^y\phi(t)dt} \quad (40)$$

In order to obtain closed-form expressions of (37)-(40), we integrate (37)-(40) by parts with respect to y between the limits 0 and ∞. Thus, we get

$$P_1(z) = P_1(0,z)\frac{\left\{1 - G_1^*\left[\omega_1\left(\lambda - \lambda C(z)\right)\right]\right\}}{\omega_1\left(\lambda - \lambda C(z)\right)} \tag{41}$$

$$P_2(z) = P_2(0,z)\frac{\left\{1 - G_2^*\left[\omega_1\left(\lambda - \lambda C(z)\right)\right]\right\}}{\omega_1\left(\lambda - \lambda C(z)\right)} \tag{42}$$

$$P_3(z) = P_3(0,z)\frac{\left\{1 - G_3^*\left[\omega_1\left(\lambda - \lambda C(z)\right)\right]\right\}}{\omega_1\left(\lambda - \lambda C(z)\right)} \tag{43}$$

$$V(z) = V(0,z)\frac{\left\{1 - W^*\left[\omega_2\left(\lambda - \lambda C(z)\right)\right]\right\}}{\omega_2\left(\lambda - \lambda C(z)\right)} \tag{44}$$

where

$$G_1^*\left[\omega_1\left(\lambda - \lambda C(z)\right)\right] = \int_0^\infty e^{-\omega_1(\lambda - \lambda C(z))y} dG_1(y) \tag{45}$$

$$G_2^*\left[\omega_1\left(\lambda - \lambda C(z)\right)\right] = \int_0^\infty e^{-\omega_1(\lambda - \lambda C(z))y} dG_2(y) \tag{46}$$

$$G_3^*\left[\omega_1\left(\lambda - \lambda C(z)\right)\right] = \int_0^\infty e^{-\omega_1(\lambda - \lambda C(z))y} dG_3(y) \tag{47}$$

$$W^*\left[\omega_2\left(\lambda - \lambda C(z)\right)\right] = \int_0^\infty e^{-\omega_2(\lambda - \lambda C(z))y} dW(y) \tag{48}$$

are the Laplace-Stieltjes transforms of type $j\left(j = 1,2,3\right)$ services and vacation time respectively.

A clear look at (37)-(40) shows that there is need to find the unknown values of $P_1(0,z)$, $P_2(0,z)$, $P_3(0,z)$, $V(0,z)$, and $R(0,z)$, respectively. So, we begin by multiplying (23) by z^{n+1}, summing both sides over n from 0 to ∞, and utilizing the probability generating functions defined in (7) together with (22) to obtain

$$\begin{aligned} zP_1(0,z) = (1-p)\xi_1\left[\int_0^\infty P_1(y,z)\mu_1(y)dy + \int_0^\infty P_2(y,z)\mu_2(y)dy + \int_0^\infty P_3(y,z)\mu_3(y)dy\right] \\ +\xi_1\int_0^\infty V(y,z)\phi(y)dy + \omega_1\xi_1\lambda\left(C(z)-1\right)Q \end{aligned} \tag{49}$$

Again, multiplying (24) by z^{n+1}, summing both sides over n from 0 to ∞, and using the probability generating functions defined in (7) together with (22), we obtain

$$\begin{aligned} zP_2(0,z) = (1-p)\xi_2\left[\int_0^\infty P_1(y,z)\mu_1(y)dy + \int_0^\infty P_2(y,z)\mu_2(y)dy + \int_0^\infty P_3(y,z)\mu_3(y)dy\right] \\ +\xi_2\int_0^\infty V(y,z)\phi(y)dy + \omega_1\xi_2\lambda\left(C(z)-1\right)Q \end{aligned} \tag{50}$$

Also, multiplying (25) by z^{n+1}, summing both sides over n from 0 to ∞, and using the probability generating functions defined in (22) together with (7), we obtain

$$zP_3(0,z)=(1-p)\xi_3\left[\int_0^\infty P_1(y,z)\mu_1(y)dy+\int_0^\infty P_2(y,z)\mu_2(y)dy+\int_0^\infty P_3(y,z)\mu_3(y)dy\right]$$
$$+\xi_3\int_0^\infty V(y,z)\phi(y)dy+\omega_1\xi_3\lambda\left(C(z)-1\right)Q \tag{51}$$

Multiplying (26) by z^{n+1}, summing both sides over n from 0 to ∞, and using the probability generating functions defined in (22) together with (7), we obtain

$$V(0,z)=p\left[\int_0^\infty P_1(y,z)\mu_1(y)dy+\int_0^\infty P_2(y,z)\mu_2(y)dy+\int_0^\infty P_3(y,z)\mu_3(y)dy\right] \tag{52}$$

To further simplify (49)-(50), we first multiply the right hand side of (37) by $\mu_1(y)$; (38) by $\mu_2(y)$; (39) by $\mu_3(y)$ and (40) by $\phi(y)$, respectively and integrating them with respect to y between the limits 0 to ∞ gives

$$\int_0^\infty P_1(y,z)\mu_1(y)dy=P_1(0,z)G_1^*\left[\omega_1\left(\lambda-\lambda C(z)\right)\right] \tag{53}$$

$$\int_0^\infty P_2(y,z)\mu_2(y)dy=P_2(0,z)G_2^*\left[\omega_1\left(\lambda-\lambda C(z)\right)\right] \tag{54}$$

$$\int_0^\infty P_3(y,z)\mu_3(y)dy=P_3(0,z)G_3^*\left[\omega_1\left(\lambda-\lambda C(z)\right)\right] \tag{55}$$

$$\int_0^\infty V(y,z)\phi(y)dy=V(0,z)W^*\left[\omega_2\left(\lambda-\lambda C(z)\right)\right] \tag{56}$$

Next, we substitute (53)-(56) into (49)-(52) noting that $\omega_1\left(\lambda-\lambda C(z)\right)=m$ and $\omega_2\left(\lambda-\lambda C(z)\right)=n$ to obtain

$$zP_1(0,z)=(1-p)\xi_1\left[P_1(0,z)G_1^*(m)+P_2(0,z)G_2^*(m)+P_3(0,z)G_3^*(m)\right]$$
$$+\xi_1 V(0,z)W^*(n)+\omega_1\xi_1\lambda\left(C(z)-1\right)Q \tag{57}$$

$$zP_2(0,z)=(1-p)\xi_2\left[P_1(0,z)G_1^*(m)+P_2(0,z)G_2^*(m)+P_3(0,z)G_3^*(m)\right]$$
$$+\xi_2 V(0,z)W^*(n)+\omega_1\xi_2\left(\lambda C(z)-\lambda\right)Q \tag{58}$$

$$zP_3(0,z)=(1-p)\xi_3\left[P_1(0,z)G_1^*(m)+P_2(0,z)G_2^*(m)+P_3(0,z)G_3^*(m)\right]$$
$$+\xi_3 V(0,z)W^*(n)+\omega_1\xi_3\left(\lambda C(z)-\lambda\right)Q \tag{59}$$

$$V(0,z)=p\left[P_1(0,z)G_1^*(m)+P_2(0,z)G_2^*(m)+P_3(0,z)G_3^*(m)\right] \tag{60}$$

Plugging (60) into (57)-(59), we obtain

$$zP_1(0,z)=(1-p)\xi_1\left[P_1(0,z)G_1^*(m)+P_2(0,z)G_2^*(m)+P_3(0,z)G_3^*(m)\right]$$
$$+\xi_1 p\left[P_1(0,z)G_1^*(m)+P_2(0,z)G_2^*(m)+P_3(0,z)G_3^*(m)\right]W^*(n) \tag{61}$$
$$+\omega_1\xi_1\left(\lambda C(z)-\lambda\right)Q$$

$$\begin{aligned} zP_2(0,z) &= (1-p)\xi_2\left[P_1(0,z)G_1^*(m)+P_2(0,z)G_2^*(m)+P_3(0,z)G_3^*(m)\right] \\ &+\xi_2 p\left[P_1(0,z)G_1^*(m)+P_2(0,z)G_2^*(m)+P_3(0,z)G_3^*(m)\right]W^*(n) \\ &+\omega_1\xi_2(\lambda C(z)-\lambda)Q \end{aligned} \quad (62)$$

$$\begin{aligned} zP_3(0,z) &= (1-p)\xi_3\left[P_1(0,z)G_1^*(m)+P_2(0,z)G_2^*(m)+P_3(0,z)G_3^*(m)\right] \\ &+\xi_3 p\left[P_1(0,z)G_1^*(m)+P_2(0,z)G_2^*(m)+P_3(0,z)G_3^*(m)\right]W^*(n) \\ &+\omega_1\xi_3(\lambda C(z)-\lambda)Q \end{aligned} \quad (63)$$

Solving (61), (62) and (63) simultaneously, we have

$$P_1(0,z)=\frac{\omega_1\xi_1(\lambda C(z)-\lambda)Q}{B(z)} \quad (64)$$

$$P_2(0,z)=\frac{\omega_1\xi_2(\lambda C(z)-\lambda)Q}{B(z)} \quad (65)$$

$$P_3(0,z)=\frac{\omega_1\xi_3(\lambda C(z)-\lambda)Q}{B(z)} \quad (66)$$

$$V(0,z)=\frac{p\,\omega_1(\lambda C(z)-\lambda)\left[\xi_1G_1^*(m)+\xi_2G_2^*(m)+\xi_3G_3^*(m)\right]Q}{B(z)} \quad (67)$$

where

$$B(z)=z-\left[(1-p)+pW^*(n)\right]\left[\xi_1P_1(0,z)G_1^*(m)+\xi_2P_2(0,z)G_2^*(m)+\xi_3P_3(0,z)G_3^*(m)\right]$$

Substituting (64) into (41); (65) into (42); (66) into (43); (67) into (44), we obtain the results shown in (28)-(31). This completes the proof of Theorem 1.

Corollary 1 *For the queueing model* $M^{[Y]}/(G_1G_2G_3)^T/1/G(BS)/V_s/B_{lk}$, *the steady-state probabilities that the server is busy providing type 1, 2 and 3 services, and the steady state probability that the server is on vacation are respectively given as follows*

$$P_1(1)=\frac{\xi_1\omega_1\lambda E(I)E(S_1)Q}{1-\lambda E(I)\left\{\omega_2pE(V)+\omega_1\left[\xi_1E(S_1)+\xi_2E(S_2)+\xi_3E(S_3)\right]\right\}} \quad (68)$$

$$P_2(1)=\frac{\xi_2\omega_1\lambda E(I)E(S_2)Q}{1-\lambda E(I)\left\{\omega_2pE(V)+\omega_1\left[\xi_1E(S_1)+\xi_2E(S_2)+\xi_3E(S_3)\right]\right\}} \quad (69)$$

$$P_3(1)=\frac{\xi_3\omega_1\lambda E(I)E(S_3)Q}{1-\lambda E(I)\left\{\omega_2pE(V)+\omega_1\left[\xi_1E(S_1)+\xi_2E(S_2)+\xi_3E(S_3)\right]\right\}} \quad (70)$$

$$V(1)=\frac{p\,\omega_1\lambda E(I)\,E(V)Q}{1-\lambda E(I)\left\{\omega_2pE(V)+\omega_1\left[\xi_1E(S_1)+\xi_2E(S_2)+\xi_3E(S_3)\right]\right\}} \quad (71)$$

where $E(I)$ is the mean size of batch of arriving customers; $E(V)$ is the mean of vacation time; $E(S_1)$, $E(S_2)$ and $E(S_3)$ are the mean service times of type 1, type 2, and type 3 services, respectively.

Proof The proof of Corollary 1 follows directly from Theorem 1 by substituting $z=1$. However, a substitution of $z=1$ in Theorem 1 shows that $P_1(z)$, $P_2(z)$, $P_3(z)$ and $V(z)$, respectively, are indeterminate of the $0/0$ form. To resolve this indeterminacy problem, we apply L'Hopital's rule on (28)-(31) respectively. In doing this, we let $m=\omega_1(\lambda-\lambda C(z))$ and $n=\omega_2(\lambda-\lambda C'(z))$, and so $\frac{dm}{dz}=-\omega_1\lambda C'(z)$, $\frac{dn}{dz}=-\omega_2\lambda C'(z)$, $\frac{dG_j^*(m)}{dz}=\frac{dG_j^*(m)}{dm}.\frac{dm}{dz}=G_j^{*\prime}(m)\left(-\omega_1\lambda C'(z)\right)$ and $\frac{dW^{*\prime}(n)}{dz}=\frac{dW^{*\prime}(n)}{dn}.\frac{dn}{dz}=W^{*\prime}(n)\left(-\omega_2\lambda C'(z)\right)$. We then obtain the first derivatives of the numerators of (28)-(31) as follows:

$$T_1'(z)=\frac{d}{dz}T_1(z)=\frac{d}{dz}\xi_1\left[1-G_1^*(m)\right]Q=\xi_1\omega_1\lambda G_1^{*\prime}(m)C'(z) \tag{72}$$

$$T_2'(z)=\frac{d}{dz}T_2(z)=\frac{d}{dz}\xi_2\left[1-G_2^*(m)\right]Q=\xi_2\omega_1\lambda G_2^{*\prime}(m)C'(z) \tag{73}$$

$$T_3'(z)=\frac{d}{dz}T_3(z)=\frac{d}{dz}\xi_3\left[1-G_3^*(m)\right]Q=\xi_3\omega_1\lambda G_3^{*\prime}(m)C'(z) \tag{74}$$

$$T_4'(z)=\frac{d}{dz}T_4(z)=\frac{d}{dz}\left\{p\frac{\omega_1}{\omega_2}\left[1-W^*(n)\right]\left[\xi_1G_1^*(m)+\xi_2G_2^*(m)+\xi_3G_3^*(m)\right]Q\right\} \tag{75}$$

$$=\frac{Qp\omega_1}{\omega_2}\left\{\begin{array}{l}\xi_1G_1^{*\prime}(m)\left(-\omega_1\lambda C'(z)\right)+\xi_2G_2^{*\prime}(m)\left(-\omega_1\lambda C'(z)\right)+\xi_3G_3^{*\prime}(m)\left(-\omega_1\lambda C'(z)\right)\\ -W^*(n)\left[\xi_1G_1^{*\prime}(m)\left(-\omega_1\lambda C'(z)\right)+\xi_2G_2^{*\prime}(m)\left(-\omega_1\lambda C'(z)\right)+\xi_3G_3^{*\prime}(m)\left(-\omega_1\lambda C'(z)\right)\right]\\ +\left[\xi_1G_1^*(m)+\xi_2G_2^*(m)+\xi_3G_3^*(m)\right]W^{*\prime}(n)\left(-\omega_2\lambda C'(z)\right)\end{array}\right\}$$

$$=\frac{Qp\omega_1}{\omega_2}\left\{\begin{array}{l}\omega_1\lambda C'(z)\left[W^*(n)-1\right]\left[\xi_1G_1^{*\prime}(m)+\xi_2G_2^{*\prime}(m)+\xi_3G_3^{*\prime}(m)\right]\\ -\omega_2\lambda C'(z)\left[\xi_1G_1^*(m)+\xi_2G_2^*(m)+\xi_3G_3^*(m)\right]W^{*\prime}(n)\end{array}\right\} \tag{76}$$

Similarly, the first derivative of (32), which serves as the denominator of (28)-(32) is as given below:

$$B'(z)=\frac{d}{dz}B(z)=\frac{d}{dz}\left\{z-\left[(1-p)+pW^*(n)\right]\left[\xi_1G_1^*(m)+\xi_2G_2^*(m)+\xi_3G_3^*(m)\right]\right\}$$

$$=1-\left\{\begin{array}{l}(1-p)\left[\xi_1 G_1^{*\prime}(m)\left(-\omega_1\lambda C'(z)\right)+\xi_2 G_2^{*\prime}(m)\left(-\omega_1\lambda C'(z)\right)+\xi_3 G_3^{*\prime}(m)\left(-\omega_1\lambda C'(z)\right)\right]\\ +pW^*(n)\left[\xi_1 G_1^{*\prime}(m)\left(-\omega_1\lambda C'(z)\right)+\xi_2 G_2^{*\prime}(m)\left(-\omega_1\lambda C'(z)\right)+\xi_3 G_3^{*\prime}(m)\left(-\omega_1\lambda C'(z)\right)\right]\\ +\left[\xi_1 G_1^*(m)+\xi_2 G_2^*(m)+\xi_3 G_3^*(m)\right]pW^{*\prime}(n)\left(-\omega_2\lambda C'(z)\right)\end{array}\right\}$$

$$=1+\left\{\begin{array}{l}\omega_1\lambda C'(z)\left[(1-p)+pW^*(n)\right]\left[\xi_1 G_1^{*\prime}(m)+\xi_2 G_2^{*\prime}(m)+\xi_3 G_3^{*\prime}(m)\right]\\ p\omega_2\lambda C'(z)\left[\xi_1 G_1^*(m)+\xi_2 G_2^*(m)+\xi_3 G_3^*(m)\right]W^{*\prime}(n)\end{array}\right\} \tag{77}$$

Evaluating the derivatives in (72)-(77) at $z=1$, and substituting $G_j^{*'}(0)=-E(S_j)$, $j=1,2,3$, $W^*(0)=1$, $W^{*'}(0)=-E(V)$ and $C^{'}(1)=E(I)$, we obtain

$$T_1'(1)=\lim_{z\to 1}T_1'(z)=\xi_1\omega_1\lambda E(I)E(S_1)Q \tag{78}$$

$$T_2'(1)=\lim_{z\to 1}T_2'(z)=\xi_2\omega_1\lambda E(I)E(S_2)Q \tag{79}$$

$$T_3'(1)=\lim_{z\to 1}T_3'(z)=\xi_3\omega_1\lambda E(I)E(S_3)Q \tag{80}$$

$$T_4'(1)=\lim_{z\to 1}T_4'(z)=p\omega_1\lambda E(I)E(V)Q \tag{81}$$

$$B'(1)=\lim_{z\to 1}B'(z)=1-\lambda E(I)\left\{\omega_2 pE(V)+\left[\xi_1 E(S_1)+\xi_2 E(S_2)+\xi_3 E(S_3)\right]\right\} \tag{82}$$

Plugging (78) and (82) into $P_1(1)=T_1'(1)/B'(1)$; (79) and (82) into $P_2(1)=T_2'(1)/B'(1)$; (80) and (82) into $P_3(1)=T_3'(1)/B'(1)$; and (81) and (82) into $P_4(1)=T_4'(1)/B'(1)$, we obtain (68)-(71) and this completes the proof of Corollary 1.

Theorem 2 *For the queueing model* $M^{[Y]}/(G_1G_2G_3)^T/1/G(BS)/V_s/B_{lk}$, *the probability generating function of the queue size irrespective of the state of the system, denoted by* $P_q(z)$, *is given by*

$$P_q(z)=\frac{Q\left[\xi_1 G_1^*(m)+\xi_2 G_2^*(m)+\xi_3 G_3^*(m)-1\right]+p\dfrac{\omega_1}{\omega_2}\left(W^*(n)-1\right)\left[\xi_1 G_1^*(m)+\xi_2 G_2^*(m)+\xi_3 G_3^*(m)\right]}{z-\left\{(1-p)+pW^*(n)\right\}\left[\xi_1 G_1^*(m)+\xi_2 G_2^*(m)+\xi_3 G_3^*(m)\right]} \tag{83}$$

Proof By adding (28), (29), (30) and (31), we have

$$P_q(z)=P_1(z)+P_2(z)+P_3(z)+V(z) \tag{84}$$

Substituting the expressions in (28)-(31) into (84), we have

$$P_q(z)=\frac{\xi_1\left[1-G_1^*(m)\right]Q}{B(z)}+\frac{\xi_2\left[1-G_2^*(m)\right]Q}{B(z)}+\frac{\xi_3\left[1-G_3^*(m)\right]Q}{B(z)}+\frac{p\dfrac{\omega_1}{\omega_2}\left[1-W^*(n)\right]\left[\xi_1 G_1^*(m)+\xi_2 G_2^*(m)+\xi_3 G_3^*(m)\right]Q}{B(z)} \tag{85}$$

On simplifying (85), we obtain (83), where $D(z)$ has been defined in (32), and this completes the proof of Theorem 2.

Corollary 2 *For the queueing model* $M^{[Y]}/(G_1G_2G_3)^T/1/G(BS)/V_s/B_{lk}$, *the steady-state probability that the server is busy, irrespective of whether it is providing type 1 or type 2 or type 3 services, and on vacation is given by:*

$$P_q(1)=\frac{\omega_1\lambda E(I)\left\{\left[\xi_1E(S_1)+\xi_2E(S_2)+\xi_3E(S_3)\right]+pE(V)\right\}Q}{1-\lambda E(I)\left\{\omega_2pE(V)+\omega_1\left[\xi_1E(S_1)+\xi_2E(S_2)+\xi_3E(S_3)\right]\right\}} \tag{86}$$

Proof By adding (69), (70), (71) and (72), we get

$$P_q(1)=P_1(1)+P_2(1)+P_3(1)+V(1) \tag{87}$$

Substituting the expressions in (68)-(71) into (87), we have

$$\begin{aligned}P_1(1)=&\frac{\xi_1\omega_1\lambda E(I)E(S_1)Q}{1-\lambda E(I)\left\{\omega_2pE(V)+\omega_1\left[\xi_1E(S_1)+\xi_2E(S_2)+\xi_3E(S_3)\right]\right\}}\\&+\frac{\xi_2\omega_1\lambda E(I)E(S_2)Q}{1-\lambda E(I)\left\{\omega_2pE(V)+\omega_1\left[\xi_1E(S_1)+\xi_2E(S_2)+\xi_3E(S_3)\right]\right\}}\\&+\frac{\xi_3\omega_1\lambda E(I)E(S_3)Q}{1-\lambda E(I)\left\{\omega_2pE(V)+\omega_1\left[\xi_1E(S_1)+\xi_2E(S_2)+\xi_3E(S_3)\right]\right\}}\\&+\frac{p\,\omega_1\,\lambda\,E(I)\,E(V)Q}{1-\lambda E(I)\left\{\omega_2pE(V)+\omega_1\left[\xi_1E(S_1)+\xi_2E(S_2)+\xi_3E(S_3)\right]\right\}}\end{aligned} \tag{88}$$

On simplifying (88), we obtain (86) and this completes the proof of Corollary 2.

4.2 Performance measures of the proposed queueing model

In this section the performance measures of the queueing system under investigation are given in Theorems 3 through 8, respectively.

Theorem 3 *For the queueing model* $M^{[Y]}/(G_1G_2G_3)^T/1/G(BS)/V_s/B_{lk}$, *the steady state probability that the server is idle, denoted by* Q, *is given by:*

$$Q=\frac{1-\lambda E(I)\left\{\omega_2pE(V)+\omega_1\left[\xi_1E(S_1)+\xi_2E(S_2)+\xi_3E(S_3)\right]\right\}}{1+p(\omega_1-\omega_2)\lambda E(I)E(V)} \tag{89}$$

Proof To prove Theorem 3, we substitute (86) into the normalizing condition $Q+P_q(1)=1$. Thus, we obtain

$$Q+\frac{\omega_1\lambda E(I)\left[\xi_1E(S_1)+\xi_2E(S_2)+\xi_3E(S_3)+pE(V)\right]Q}{1-\lambda E(I)\left\{\omega_2pE(V)+\omega_1\left[\xi_1E(S_1)+\xi_2E(S_2)+\xi_3E(S_3)\right]\right\}}=1 \tag{90}$$

From (90), we make Q subject of the formula to get (89) and this completes the proof of Theorem 3.

Theorem 4 *For the queueing model* $M^{[Y]}/(G_1G_2G_3)^T/1/G(BS)/V_s/B_{lk}$, *the probability that the server is busy (utilization factor), denoted by* ρ, *is given by:*

$$\rho=\frac{\omega_1\lambda E(I)\left[pE(V)+\left\{\xi_1E(S_1)+\xi_2E(S_2)+\xi_3E(S_3)\right\}\right]}{1+p(\omega_1-\omega_2)\lambda E(I)E(V)} \tag{91}$$

where $\rho<1$ is the stability condition under which the steady state exists.

Proof To prove Theorem 4, we substitute (89) into the relation $Q+\rho=1$. Thus, we obtain

$$\frac{1-\lambda E(I)\left\{\omega_2pE(V)+\omega_1\left[\xi_1E(S_1)+\xi_2E(S_2)+\xi_3E(S_3)\right]\right\}}{1+p(\omega_1-\omega_2)\lambda E(I)E(V)}+\rho=1 \tag{92}$$

On making ρ subject of the formula from (92), we obtain the result shown in (91) and this completes the proof of Theorem 4.

Theorem 5 *For the queueing model* $M^{[Y]}/(G_1G_2G_3)^T/1/G(BS)/V_s/B_{lk}$, *the average queue size at random epoch, denoted by* L_q, *is given by*

$$L_q=\frac{B'(1)T''(1)-T'(1)B''(1)}{2\left(B'(1)\right)^2} \tag{93}$$

where

$$T'(1)=\omega_1\lambda E(I)\left\{\left[\xi_1E(S_1)+\xi_2E(S_2)+\xi_3E(S_3)\right]+pE(V)\right\}Q \tag{94}$$

$$T''(1)=Q\begin{bmatrix}\omega_1\lambda E(I(I-1))\left\{\left[\xi_1E(S_1)+\xi_2E(S_2)+\xi_3E(S_3)\right]+pE(V)\right\}\\+\left(\omega_1\lambda E(I)\right)^2\left\{\left[\xi_1E(S_1^2)+\xi_2E(S_2^2)+\xi_3E(S_3^2)\right]+2pE(V)\left[\xi_1E(S_1)+\xi_2E(S_2)+\xi_3E(S_3)\right]\right\}\\+p\omega_1\omega_2\left(\lambda E(I)\right)^2E(V^2)\end{bmatrix} \tag{95}$$

$$B'(1)=1-\lambda E(I)\left\{\left[\omega_1\xi_1E(S_1)+\omega_1\xi_2E(S_2)+\omega_1\xi_3E(S_3)\right]+p\omega_2E(V)\right\} \tag{96}$$

$$\begin{aligned}B''(1)=&-\left(\lambda\omega_1E(I)\right)^2\left[\xi_1E(S_1^2)+\xi_2E(S_2^2)+\xi_3E(S_3^2)\right]-p\left(\omega_2\lambda E(I)\right)^2E(V^2)\\&-\lambda E(I(I-1))\left\{\left[\omega_1\xi_1E(S_1)+\omega_1\xi_2E(S_2)+\omega_1\xi_3E(S_3)\right]+p\omega_2E(V)\right\}\\&-2p\omega_1\omega_2\left(\lambda E(I)\right)^2E(V)\left[\xi_1E(S_1)+\xi_2E(S_2)+\xi_3E(S_3)\right]\end{aligned} \tag{97}$$

and $E(S_1^2)$, $E(S_2^2)$, $E(S_3^2)$ and $E(V^2)$ are the second moments of service times type 1, type 2, type 3 and vacation time respectively; $E(I(I-1))$ is the second factorial moment of the batch of arriving customers and Q is given in (89).

Proof Let $T(z)$ and $B(z)$ denote the top and bottom of the right hand side of (83), then

$$T(z)=Q\left[1-\xi_1 G_1^*(m)+\xi_2 G_2^*(m)+\xi_3 G_3^*(m)\right]+p\frac{\omega_1}{\omega_2}\left(1-W^*(n)\right)\left[\xi_1 G_1^*(m)+\xi_2 G_2^*(m)+\xi_3 G_3^*(m)\right]$$

(98)

$$B(z)=z-\left\{(1-p)+pW^*(n)\right\}\left[\xi_1 G_1^*(m)+\xi_2 G_2^*(m)+\xi_3 G_3^*(m)\right] \quad (99)$$

Consequently, (83) becomes

$$P_q(z)=\frac{T(z)}{B(z)} \quad (100)$$

According to Kashyap and Chaudhry [22], the application of quotient rule twice on (100) and evaluating at $z=1$, gives

$$L_q=\frac{d}{dz}P_q(z)\Big|_{z=1}=\lim_{z\to 1}\frac{T(z)}{B(z)}=\frac{B'(1)T''(1)-T'(1)B''(1)}{2\left(B'(1)\right)^2}$$

In what follows, we need to obtain the expressions for $T'(z)$, $T''(z)$, $B'(z)$ and $B''(z)$ respectively. Carrying out the first and second derivatives on (98) with respect to z, we obtain

$$T'(z)=\frac{d}{dz}T(z)=Q\frac{d}{dz}\left\{\begin{array}{l}\left[\xi_1 G_1^*(m)+\xi_2 G_2^*(m)+\xi_3 G_3^*(m)-1\right]\\ +p\frac{\omega_1}{\omega_2}\left(W^*(n)-1\right)\left[\xi_1 G_1^*(m)+\xi_2 G_2^*(m)+\xi_3 G_3^*(m)\right]\end{array}\right\}$$

$$=Q\left\{\begin{array}{l}\left[\xi_1 G_1^{*\prime}(m)\left(-\omega_1\lambda C'(z)\right)+\xi_2 G_2^{*\prime}(m)\left(-\omega_1\lambda C'(z)\right)+\xi_3 G_3^{*\prime}(m)\left(-\omega_1\lambda C'(z)\right)\right]\\ +p\frac{\omega_1}{\omega_2}W^*(n)\left[\xi_1 G_1^{*\prime}(m)\left(-\omega_1\lambda C'(z)\right)+\xi_2 G_2^{*\prime}(m)\left(-\omega_1\lambda C'(z)\right)+\xi_3 G_3^{*\prime}(m)\left(-\omega_1\lambda C'(z)\right)\right]\\ +p\frac{\omega_1}{\omega_2}\left[\xi_1 G_1^*(m)+\xi_2 G_2^*(m)+\xi_3 G_3^*(m)\right]W^{*\prime}(n)\left(-\omega_2\lambda C'(z)\right)\\ -p\frac{\omega_1}{\omega_2}\left[\xi_1 G_1^{*\prime}(m)\left(-\omega_1\lambda C'(z)\right)+\xi_2 G_2^{*\prime}(m)\left(-\omega_1\lambda C'(z)\right)+\xi_3 G_3^{*\prime}(m)\left(-\omega_1\lambda C'(z)\right)\right]\end{array}\right\}$$

$$=Q\left\{\begin{array}{l}\omega_1\lambda C'(z)\left[p\frac{\omega_1}{\omega_2}-p\frac{\omega_1}{\omega_2}W^*(n)-1\right]\left[\xi_1 G_1^{*\prime}(m)+\xi_2 G_2^{*\prime}(m)+\xi_3 G_3^{*\prime}(m)\right]\\ +p\omega_1\lambda C'(z)\left[\xi_1 G_1^*(m)+\xi_2 G_2^*(m)+\xi_3 G_3^*(m)\right]W^{*\prime}(n)\end{array}\right\} \quad (101)$$

$$T''(z)=Q\frac{d}{dz}\left\{\begin{array}{l}\omega_1\lambda C(z)\left[p\frac{\omega_1}{\omega_2}-p\frac{\omega_1}{\omega_2}W^*(n)-1\right]\left[\xi_1 G_1^{*\prime}(m)+\xi_2 G_2^{*\prime}(m)+\xi_3 G_3^{*\prime}(m)\right]\\ +p\omega_1\lambda C'(z)\left[\xi_1 G_1^*(m)+\xi_2 G_2^*(m)+\xi_3 G_3^*(m)\right]W^{*\prime}(n)\end{array}\right\}$$

$$=\left\{\begin{array}{l}\left(p\frac{\omega_1}{\omega_2}-p\frac{\omega_1}{\omega_2}W^*(n)-1\right)\left\{\begin{array}{l}\omega_1\lambda C''(z)\left[\xi_1 G_1^{*\prime}(m)+\xi_2 G_2^{*\prime}(m)+\xi_3 G_3^{*\prime}(m)\right]\\-\left(\omega_1\lambda C'(z)\right)^2\left[\xi_1 G_1^{*\prime\prime}(m)+\xi_2 G_2^{*\prime\prime}(m)+\xi_3 G_3^{*\prime\prime}(m)\right]\end{array}\right\}\\+2p\left(\omega_1\lambda C'(z)\right)^2\left[\xi_1 G_1^{*\prime}(m)+\xi_2 G_2^{*\prime}(m)+\xi_3 G_3^{*\prime}(m)\right]W^{*\prime}(n)\\+p\omega_1\left[\lambda C''(z)W^{*\prime}(n)+\omega_2\left(\lambda C'(z)\right)^2 W^{*\prime\prime}(n)\right]\left[\xi_1 G_1^*(m)+\xi_2 G_2^*(m)+\xi_3 G_3^*(m)\right]\end{array}\right\} \quad (102)$$

Similarly, the first and second derivatives of (99) with respect to z are given as follows:

$$B'(z)=\frac{d}{dz}B(z)=\frac{d}{dz}\left\{z-\left[(1-p)+pW^*(n)\right]\left[\xi_1 G_1^*(m)+\xi_2 G_2^*(m)+\xi_3 G_3^*(m)\right]\right\}$$

$$\begin{aligned}&=1+\omega_1\lambda C'(z)\left[(1-p)+pW^*(n)\right]\left[\xi_1 G_1^{*\prime}(m)+\xi_2 G_2^{*\prime}(m)+G_3^{*\prime}(m)\right]\\&\quad+p\omega_2\lambda C'(z)\left[\xi_1 G_1^*(m)+\xi_2 G_2^*(m)+\xi_3 G_3^*(m)\right]W^{*\prime}(n)\end{aligned} \quad (103)$$

$$B''(z)=\frac{d}{dz}\left\{\begin{array}{l}1+\omega_1\lambda C'(z)\left[(1-p)+pW^*(n)\right]\left[\xi_1 G_1^{*\prime}(m)+\xi_2 G_2^{*\prime}(m)+G_3^{*\prime}(m)\right]\\+p\omega_2\lambda C'(z)\left[\xi_1 G_1^*(m)+\xi_2 G_2^*(m)+\xi_3 G_3^*(m)\right]W^{*\prime}(n)\end{array}\right\}$$

$$=\left\{\begin{array}{l}\left[(1-p)+pW^*(n)\right]\left\{\begin{array}{l}\omega_1\lambda C''(z)\left[\xi_1 G_1^{*\prime}(m)+\xi_2 G_2^{*\prime}(m)+\xi_3 G_3^{*\prime}(m)\right]\\-\left(\omega_1\lambda C'(z)\right)^2\left[\xi_1 G_1^{*\prime\prime}(m)+\xi_2 G_2^{*\prime\prime}(m)+\xi_3 G_3^{*\prime\prime}(m)\right]\end{array}\right\}\\-2p\omega_1\omega_2\left(\lambda C'(z)\right)^2\left[\xi_1 G_1^{*\prime}(m)+\xi_2 G_2^{*\prime}(m)+\xi_3 G_3^{*\prime}(m)\right]W^{*\prime}(n)\\+p\omega_2\left[\lambda C''(z)W^{*\prime}(n)-\left(\omega_2\lambda C'(z)\right)^2 W^{*\prime\prime}(n)\right]\left[\xi_1 G_1^*(m)+\xi_2 G_2^*(m)+\xi_3 G_3^*(m)\right]\end{array}\right\} \quad (104)$$

Evaluating the derivatives in (101)-(104) at $z=1$, and substituting $W^*(0)=1$, $W^{*\prime}(0)=-E(V)$, $W^{*\prime\prime}(0)=E(V^2)$ $G_j^*(0)=1,\ j=1,2,3$, $G_j^{*\prime}(0)=-E(S_j),\ j=1,2,3$, $G_j^{*\prime\prime}(0)=E(S_j^2),\ j=1,2,3$, $C'(1)=E(I)$, and $C''(1)=E(I(I-1))$ we obtain (94)-(97). On putting the results in (94)-(97) into (93), we obtain a closed form expression for L_q.

Theorem 6 *For the queueing model* $M^{[Y]}/(G_1G_2G_3)^T/1/G(BS)/V_s/B_{lk}$, *the average waiting time in the queue at random epoch, denoted by* W_q, *is given by*

$$W_q = \lambda^{-1}\left[\frac{B'(1)T''(1)-T'(1)B''(1)}{2(B'(1))^2}\right] \tag{105}$$

Proof The proof of Theorem 6 follows directly from Theorem 5 by substituting (93) into Little's formula $W_q = L_q/\lambda$.This completes the proof of Theorem 6.

Theorem 7 *For the queueing model* $M^{[Y]}/(G_1G_2G_3)^T/1/G(BS)/V_s/B_{lk}$, *the average size of the queuing system at random epoch, denoted by* L_s, *is given by*

$$L_s = \frac{B'(1)T''(1)-T'(1)B''(1)}{2(B'(1))^2} + \frac{\omega_1\lambda E(I)\left[pE(V)+\{\xi_1E(S_1)+\xi_2E(S_2)+\xi_3E(S_3)\}\right]}{1+p(\omega_1-\omega_2)\lambda E(I)E(V)} \tag{106}$$

Proof Substituting (91) and (93) into Little's formula $L_s = L_q + \rho$, we have

$$L_s = \frac{B'(1)T''(1)-T'(1)B''(1)}{2(B'(1))^2} + \frac{\omega_1\lambda E(I)\left[pE(V)+\{\xi_1E(S_1)+\xi_2E(S_2)+\xi_3E(S_3)\}\right]}{1+p(\omega_1-\omega_2)\lambda E(I)E(V)} \tag{107}$$

Simplifying (107), we obtain (106) and this completes the proof of Theorem 7.

Theorem 8 *For the queueing model* $M^{[Y]}/(G_1G_2G_3)^T/1/G(BS)/V_s/B_{lk}$, *the mean waiting time in the system at random epoch, denoted by* W_s, *is given by*

$$W_s = \frac{\left[B'(z)T''(z)-T'(z)B''(z)\right]\left[1+p(\omega_1-\omega_2)\lambda E(I)E(V)\right]}{\lambda(B'(z))^2\left[1+p(\omega_1-\omega_2)\lambda E(I)E(V)\right]} + \frac{\omega_1\lambda E(I)\left[pE(V)+\{\xi_1E(S_1)+\xi_2E(S_2)+\xi_3E(S_3)\}\right](B^1(z))^2}{\lambda(B^1(z))^2\left[1+p(\omega_1-\omega_2)\lambda E(I)E(V)\right]} \tag{108}$$

Proof Substituting (93) into Little's formula $W_s = L_q/\lambda$, we have

$$W_s = \frac{1}{\lambda}\left[\frac{B'(1)T''(1)-T'(1)B''(1)}{2(B'(1))^2} + \frac{\omega_1\lambda E(I)\left[pE(V)+\{\xi_1E(S_1)+\xi_2E(S_2)+\xi_3E(S_3)\}\right]}{1+p(\omega_1-\omega_2)\lambda E(I)E(V)}\right] \tag{109}$$

On simplifying (109), we get (108). This completes the proof of Theorem 8.

5 Special cases of the batch arrival queueing system with balking, three types of heterogeneous service and Bernoulli schedule server vacation

Some queueing models are found to be sub-models of the proposed queuing model. For instance consider the following cases:

Case 1 (*No Balking*) In this case we consider a situation where there is no balking i.e. all customers join the system during busy and vacation periods. So, we let the balking parameters $\omega_1 = \omega_2 = 1$. Then our model reduces to a batch arrival queueing model with Bernoulli vacation and three types of heterogeneous service.

$$P_q(z) = \frac{Q\left[\xi_1 G_1^*(m) + \xi_2 G_2^*(m) + \xi_3 G_3^*(m) - 1\right] + p\left(W^*(n) - 1\right)\left[\xi_1 G_1^*(m) + \xi_2 G_2^*(m) + \xi_3 G_3^*(m)\right]}{z - \left\{(1-p) + pW^*(n)\right\}\left[\xi_1 G_1^*(m) + \xi_2 G_2^*(m) + \xi_3 G_3^*(m)\right]} \tag{110}$$

$$P_1(1) = \frac{\xi_1 \lambda E(I) E(S_1) Q}{1 - \lambda E(I)\left\{pE(V) + \left[\xi_1 E(S_1) + \xi_2 E(S_2) + \xi_3 E(S_3)\right]\right\}} \tag{111}$$

$$P_2(1) = \frac{\xi_2 \lambda E(I) E(S_2) Q}{1 - \lambda E(I)\left\{pE(V) + \left[\xi_1 E(S_1) + \xi_2 E(S_2) + \xi_3 E(S_3)\right]\right\}} \tag{112}$$

$$P_3(1) = \frac{\xi_3 \lambda E(I) E(S_3) Q}{1 - \lambda E(I)\left\{pE(V) + \left[\xi_1 E(S_1) + \xi_2 E(S_2) + \xi_3 E(S_3)\right]\right\}} \tag{113}$$

$$V(1) = \frac{p \lambda E(I)\, E(V) Q}{1 - \lambda E(I)\left\{pE(V) + \left[\xi_1 E(S_1) + \xi_2 E(S_2) + \xi_3 E(S_3)\right]\right\}} \tag{114}$$

$$P_q(1) = \frac{\lambda E(I)\left\{\left[\xi_1 E(S_1) + \xi_2 E(S_2) + \xi_3 E(S_3)\right] + pE(V)\right\} Q}{1 - \lambda E(I)\left\{pE(V) + \left[\xi_1 E(S_1) + \xi_2 E(S_2) + \xi_3 E(S_3)\right]\right\}} \tag{115}$$

$$Q = 1 - \lambda E(I)\left\{pE(V) + \left[\xi_1 E(S_1) + \xi_2 E(S_2) + \xi_3 E(S_3)\right]\right\} \tag{116}$$

$$\rho = \lambda E(I)\left[pE(V) + \left\{\xi_1 E(S_1) + \xi_2 E(S_2) + \xi_3 E(S_3)\right\}\right] \tag{117}$$

$$T'(1) = \lambda E(I)\left\{\left[\xi_1 E(S_1) + \xi_2 E(S_2) + \xi_3 E(S_3)\right] + pE(V)\right\} Q \tag{118}$$

$$T''(1) = Q\begin{bmatrix} \lambda E\left(I(I-1)\right)\left\{\left[\xi_1 E(S_1) + \xi_2 E(S_2) + \xi_3 E(S_3)\right] + pE(V)\right\} \\ +\left(\lambda E(I)\right)^2\left\{\left[\xi_1 E(S_1^2) + \xi_2 E(S_2^2) + \xi_3 E(S_3^2)\right] + 2pE(V)\left[\xi_1 E(S_1) + \xi_2 E(S_2) + \xi_3 E(S_3)\right]\right\} \\ +p\left(\lambda E(I)\right)^2 E(V^2) \end{bmatrix} \tag{119}$$

$$B'(1) = 1 - \lambda E(I)\left\{\left[\xi_1 E(S_1) + \xi_2 E(S_2) + \xi_3 E(S_3)\right] + pE(V)\right\} \tag{120}$$

$$B''(1) = -(\lambda E(I))^2 \left[\xi_1 E(S_1^2) + \xi_2 E(S_2^2) + \xi_3 E(S_3^2)\right] - p(\lambda E(I))^2 E(V^2) \\ -\lambda E(I(I-1))\left\{\left[\xi_1 E(S_1) + \xi_2 E(S_2) + \xi_3 E(S_3)\right] + pE(V)\right\} \\ -2p(\lambda E(I))^2 E(V)\left[\xi_1 E(S_1) + \xi_2 E(S_2) + \xi_3 E(S_3)\right] \quad (121)$$

With (118)-(121), we can obtain the average number of customers in the queue (L_q) using (89). Moreover, the values of average waiting time of a customer in the queue (W_q), average waiting time of a customer in the system (W_q), and average number of customers in the system (L_s) are obtained using the relations $W_q = L_q/\lambda$, $W_s = L_q/\lambda$ and $L_s = L_q + \rho$ (see Little, 1961) respectively.

Case 2 ***(No Vacation)*** In this case we consider a situation where the server does not take vacation. So, we let the vacation parameter $p = 0$. Then our model reduces batch arrival queueing model with balking and three types of heterogeneous service.

$$P_q(z) = \frac{Q\left[\xi_1 G_1^*(m) + \xi_2 G_2^*(m) + \xi_3 G_3^*(m) - 1\right]}{z - \left[\xi_1 G_1^*(m) + \xi_2 G_2^*(m) + \xi_3 G_3^*(m)\right]} \quad (122)$$

$$P_1(1) = \frac{\xi_1 \omega_1 \lambda E(I) E(S_1) Q}{1 - \lambda E(I)\omega_1\left[\xi_1 E(S_1) + \xi_2 E(S_2) + \xi_3 E(S_3)\right]} \quad (123)$$

$$P_2(1) = \frac{\xi_2 \omega_1 \lambda E(I) E(S_2) Q}{1 - \lambda E(I)\omega_1\left[\xi_1 E(S_1) + \xi_2 E(S_2) + \xi_3 E(S_3)\right]} \quad (124)$$

$$P_3(1) = \frac{\xi_3 \omega_1 \lambda E(I) E(S_3) Q}{1 - \lambda E(I)\omega_1\left[\xi_1 E(S_1) + \xi_2 E(S_2) + \xi_3 E(S_3)\right]} \quad (125)$$

$$V(1) = \frac{p\,\omega_1 \lambda E(I) E(V) Q}{1 - \lambda E(I)\omega_1\left[\xi_1 E(S_1) + \xi_2 E(S_2) + \xi_3 E(S_3)\right]} \quad (126)$$

$$P_q(1) = \frac{\omega_1 \lambda E(I)\left[\xi_1 E(S_1) + \xi_2 E(S_2) + \xi_3 E(S_3)\right] Q}{1 - \lambda E(I)\omega_1\left[\xi_1 E(S_1) + \xi_2 E(S_2) + \xi_3 E(S_3)\right]} \quad (127)$$

$$Q = 1 - \lambda E(I)\omega_1\left[\xi_1 E(S_1) + \xi_2 E(S_2) + \xi_3 E(S_3)\right] \quad (128)$$

$$\rho = \omega_1 \lambda E(I)\left[\left\{\xi_1 E(S_1) + \xi_2 E(S_2) + \xi_3 E(S_3)\right\}\right] \quad (129)$$

$$T'(1) = \omega_1 \lambda E(I)\left[\xi_1 E(S_1) + \xi_2 E(S_2) + \xi_3 E(S_3)\right] Q \quad (130)$$

$$T''(1) = Q\left[\begin{array}{l} \omega_1 \lambda E(I(I-1))\left[\xi_1 E(S_1) + \xi_2 E(S_2) + \xi_3 E(S_3)\right] \\ +(\omega_1 \lambda E(I))^2 \left\{\left[\xi_1 E(S_1^2) + \xi_2 E(S_2^2) + \xi_3 E(S_3^2)\right]\right\} \end{array}\right] \quad (131)$$

$$B'(1) = 1 - \lambda E(I)\left[\omega_1 \xi_1 E(S_1) + \omega_1 \xi_2 E(S_2) + \omega_1 \xi_3 E(S_3)\right] \quad (132)$$

$$B''(1) = -\left(\lambda\omega_1 E(I)\right)^2 \left[\xi_1 E\left(S_1^2\right) + \xi_2 E\left(S_2^2\right) + \xi_3 E\left(S_3^2\right)\right] \\ -\lambda E\left(I(I-1)\right)\left[\omega_1\xi_1 E(S_1) + \omega_1\xi_2 E(S_2) + \omega_1\xi_3 E(S_3)\right] \quad (133)$$

The performance measures can be obtained in the same manner as in Case 1.

Case 3 ***(No Vacation and No Balking)*** In this case we consider a situation where the server does not take vacation and there is no balking. So, we let the vacation parameter $p = 0$ and the balking parameters $\omega_1 = \omega_2 = 1$. Then our model reduces batch arrival queueing model with three types of heterogeneous service.

$$P_q(z) = \frac{Q\left[\xi_1 G_1^*(m) + \xi_2 G_2^*(m) + \xi_3 G_3^*(m) - 1\right]}{z - \left[\xi_1 G_1^*(m) + \xi_2 G_2^*(m) + \xi_3 G_3^*(m)\right]} \quad (134)$$

$$P_1(1) = \frac{\xi_1 \lambda E(I) E(S_1) Q}{1 - \lambda E(I)\left[\xi_1 E(S_1) + \xi_2 E(S_2) + \xi_3 E(S_3)\right]} \quad (135)$$

$$P_2(1) = \frac{\xi_2 \lambda E(I) E(S_2) Q}{1 - \lambda E(I)\left[\xi_1 E(S_1) + \xi_2 E(S_2) + \xi_3 E(S_3)\right]} \quad (136)$$

$$P_3(1) = \frac{\xi_3 \lambda E(I) E(S_3) Q}{1 - \lambda E(I)\left[\xi_1 E(S_1) + \xi_2 E(S_2) + \xi_3 E(S_3)\right]} \quad (137)$$

$$V(1) = \frac{p\,\lambda E(I)\,E(V) Q}{1 - \lambda E(I)\left[\xi_1 E(S_1) + \xi_2 E(S_2) + \xi_3 E(S_3)\right]} \quad (138)$$

$$P_q(1) = \frac{\lambda E(I)\left[\xi_1 E(S_1) + \xi_2 E(S_2) + \xi_3 E(S_3)\right] Q}{1 - \lambda E(I)\left[\xi_1 E(S_1) + \xi_2 E(S_2) + \xi_3 E(S_3)\right]} \quad (139)$$

$$Q = 1 - \lambda E(I)\left[\xi_1 E(S_1) + \xi_2 E(S_2) + \xi_3 E(S_3)\right] \quad (140)$$

$$\rho = \lambda E(I)\left[\left\{\xi_1 E(S_1) + \xi_2 E(S_2) + \xi_3 E(S_3)\right\}\right] \quad (141)$$

$$T'(1) = \lambda E(I)\left[\xi_1 E(S_1) + \xi_2 E(S_2) + \xi_3 E(S_3)\right] Q \quad (142)$$

$$T''(1) = Q\begin{bmatrix} \lambda E\left(I(I-1)\right)\left[\xi_1 E(S_1) + \xi_2 E(S_2) + \xi_3 E(S_3)\right] \\ +\left(\lambda E(I)\right)^2 \left\{\left[\xi_1 E\left(S_1^2\right) + \xi_2 E\left(S_2^2\right) + \xi_3 E\left(S_3^2\right)\right]\right\} \end{bmatrix} \quad (143)$$

$$B'(1) = 1 - \lambda E(I)\left[\xi_1 E(S_1) + \xi_2 E(S_2) + \xi_3 E(S_3)\right] \quad (144)$$

$$B''(1) = -\left(\lambda E(I)\right)^2 \left[\xi_1 E\left(S_1^2\right) + \xi_2 E\left(S_2^2\right) + \xi_3 E\left(S_3^2\right)\right] \\ -\lambda E\left(I(I-1)\right)\left[\xi_1 E(S_1) + \xi_2 E(S_2) + \xi_3 E(S_3)\right] \quad (145)$$

Case 4 (***No third type of service)*** In this case we consider a situation whereby the server is providing only two types of service and with balking and Bernoulli schedule server vacation. We then let the parameter of type 3 service to be zero i.e., $\xi_3 = 0$ and $\xi_1 + \xi_2 = 1$. In consequence, our model reduces batch arrival queueing model with three types of heterogeneous service, balking and Bernoulli schedule server vacation. This type of model was studied by Baruah et al. (2013).

$$P_q(z) = \frac{Q\left[\xi_1 G_1^*(m) + \xi_2 G_2^*(m) - 1\right] + p\frac{\omega_1}{\omega_2}\left(W^*(n) - 1\right)\left[\xi_1 G_1^*(m) + \xi_2 G_2^*(m)\right]}{z - \left\{(1-p) + pW^*(n)\right\}\left[\xi_1 G_1^*(m) + \xi_2 G_2^*(m)\right]} \tag{146}$$

$$P_1(1) = \frac{\xi_1\omega_1\lambda E(I)E(S_1)Q}{1 - \lambda E(I)\left\{\omega_2 pE(V) + \omega_1\left[\xi_1 E(S_1) + \xi_2 E(S_2)\right]\right\}} \tag{147}$$

$$P_2(1) = \frac{\xi_2\omega_1\lambda E(I)E(S_2)Q}{1 - \lambda E(I)\left\{\omega_2 pE(V) + \omega_1\left[\xi_1 E(S_1) + \xi_2 E(S_2)\right]\right\}} \tag{148}$$

$$V(1) = \frac{p\,\omega_1\,\lambda\,E(I)\,E(V)Q}{1 - \lambda E(I)\left\{\omega_2 pE(V) + \omega_1\left[\xi_1 E(S_1) + \xi_2 E(S_2)\right]\right\}} \tag{149}$$

$$P_q(1) = \frac{\omega_1\lambda E(I)\left\{\left[\xi_1 E(S_1) + \xi_2 E(S_2)\right] + pE(V)\right\}Q}{1 - \lambda E(I)\left\{\omega_2 pE(V) + \omega_1\left[\xi_1 E(S_1) + \xi_2 E(S_2)\right]\right\}} \tag{150}$$

$$Q = \frac{1 - \lambda E(I)\left\{\omega_2 pE(V) + \omega_1\left[\xi_1 E(S_1) + \xi_2 E(S_2)\right]\right\}}{1 + p(\omega_1 - \omega_2)\lambda E(I)E(V)} \tag{151}$$

$$\rho = \frac{\omega_1\lambda E(I)\left[pE(V) + \left\{\xi_1 E(S_1) + \xi_2 E(S_2)\right\}\right]}{1 + p(\omega_1 - \omega_2)\lambda E(I)E(V)} \tag{152}$$

$$T'(1) = \omega_1\lambda E(I)\left\{\left[\xi_1 E(S_1) + \xi_2 E(S_2)\right] + pE(V)\right\}Q \tag{153}$$

$$T''(1) = Q\begin{bmatrix} \omega_1\lambda E\left(I(I-1)\right)\left\{\left[\xi_1 E(S_1) + \xi_2 E(S_2)\right] + pE(V)\right\} \\ + \left(\omega_1\lambda E(I)\right)^2\left\{\left[\xi_1 E(S_1^2) + \xi_2 E(S_2^2)\right] + 2pE(V)\left[\xi_1 E(S_1) + \xi_2 E(S_2)\right]\right\} \\ + p\omega_1\omega_2\left(\lambda E(I)\right)^2 E(V^2) \end{bmatrix} \tag{154}$$

$$B'(1) = 1 - \lambda E(I)\left\{\left[\omega_1\xi_1 E(S_1) + \omega_1\xi_2 E(S_2)\right] + p\omega_2 E(V)\right\} \tag{155}$$

$$\begin{aligned} B''(1) = &-\left(\lambda\omega_1 E(I)\right)^2\left[\xi_1 E(S_1^2) + \xi_2 E(S_2^2)\right] - p\left(\omega_2\lambda E(I)\right)^2 E(V^2) \\ &-\lambda E\left(I(I-1)\right)\left\{\left[\omega_1\xi_1 E(S_1) + \omega_1\xi_2 E(S_2)\right] + p\omega_2 E(V)\right\} \\ &-2p\omega_1\omega_2\left(\lambda E(I)\right)^2 E(V)\left[\xi_1 E(S_1) + \xi_2 E(S_2)\right] \end{aligned} \tag{156}$$

6 Conclusion

Presented in this paper is a non-Markovian queue with three types of service, balking and Bernoulli server vacation. The elapsed service time and vacation time have been introduced as supplementary variables and the model is described as a bivariate Markov process. Consequently, the non-Markovian process is converted to Markovian process. The steady state probability generating function of the queue size has been derived. Also, the steady state system performance measures like the average number of customers in the queue and in the system, the average waiting time in the queue and in the system, the idle state probability, the utilization factor are obtained. We have also discussed some special cases of the proposed to show that the results found in literature, in particular Baruah et al. (2012, 2013), coincide with our results.

Acknowledgments The authors would like to thank the reviewers for their valuable comments and suggestions, which helped to improve the presentation of this paper.

Conflict of interest: There is no conflict of interest between the authors

References

Altman, E., Yechiali, U.: Analysis of customers' impatience in queue with server vacations. *Queuing Syst.* 52(4), 261-279 (2006)

Anabosi, R.F., Madan, K.C.: A single server queue with two types of service, Bernoulli schedule server vacations and a single vacation policy. *Pakistan J. Stat.* 19(3), 331-342 (2003)

Baba, Y.: On the $M^X/G/1$ queue with vacation time. *Oper. Res. Lett.* 5(2), 93–98 (1986)

Bailey, N.T.J.: On queuing processes with bulk service. *J. Royal Stat. Soc.* B 16: 80-97 (1954)

Barrer D.Y.: Queuing with impatient customers and ordered service. *Oper. Res.* 5, 656-650 (1957)

Baruah, M., Madan, K.C. and Eldabi, T.: Balking and Re-service in a Vacation Queue with Batch Arrival and Two Types of Heterogeneous Service. *J. Math. Res. 4(4), 114-124* (2012)

Baruah, M., Madan, K.C., Eldabi, T.: Balking and re-service in a vacation queue with batch arrival and two types of heterogeneous service. *J. Math. Res.* 4(4), 114-124 (2012)

Borthakur, A., Choudhury, G.: On a batch arrival Poisson queue with generalized vacation. Sankhya: *The Indian J. Stat.* 59(3), 369-383 (1997)

Borthakur, A., Medhi, J.: A queueing system with arrival and service in batches of variable size. Cah. Du. centre d'Et.de Rech. *Oper. Res., 16, 117-126* (1974)

Chao, X., Zhao, Y.Q.: Analysis of multi-server queues with station and server vacations. *Euro. J. Oper. Res.*, 110(2), 392-406 (1998)

Choi, B.D., Park, K.K.: The $M/G/1$ retrial queue with Bernoulli schedule. *Queueing Syst.* 7, 219-228 (1990)

Choudhury, G., Mandhuchanda, P.: A two phase queueing system with Bernoulli feedback. *Inf. Manage Sci.* 16(1), 35-52 (2005)

Choudhury, G.: An $M^X/G/1$ queueing system with a setup period and a vacation period. *Queueing Syst.* 36(1), 23-38 (2000)

Choudhury, G.: Analysis of the $M^{[X]}/G/1$ queueing system with vacation times. *Sankhyá Series* -B 64 (1), 37-49 (2002)

Cohen, J.W.: *The single server queue*, second ed., Amsterdam, North-Holland, The Netherlands (1982)

Cox, D.R.: The analysis of non-Markovian stochastic process by the inclusion of supplementary variables. *Proc. Camb. Phil. Soc.* 51, 433-441 (1955)

Cramer, M.: Stationary distributions in a queueing system with vacation times and limited service. *Queueing Syst.* 4(1), 57-68 (1989)

Doshi, B.T.: Queueing systems with vacations, a survey. *Queueing Syst.* 1(1), 29-66 (1986)

Doshi, B.T.: Analysis of a two phase queueing system with general service times. *Opera Res. Lett.* 10, 275-265 (1991)

Ebenesar Anna Bagyam J., Udaya Chandrika, K.: Bulk arrival two phase retrial queue with two types service and extended bernoulli vacation. *Intl. J. Math. Trends. and Tech.* 4(7), 116-124 (2013)

Ebenesar Anna Bagyam, J., Udaya Chandrika, K.: Non-persistent retrial queueing system with two types of heterogeneous service. *Global J. Theo. and Appl. Math. Sciences* 1(2), 157-164 (2011)

Enogwe, S.U., Obiora-Ilouno, H.O.: Effects of Reneging, Server Breakdowns and Vacation on a Batch Arrival Single Server Queueing System with Three Fluctuating Modes of Service. *Open J. Opt.* 9, 105-128 (2020)

Enogwe, S.U., Onyeagu, S.I., Obiora-Ilouno, H.O.: Single Channel Batch Arrival Queueing Model for Systems that Provides Three-Stage Service for Customers that Renege During Server Vacation and Breakdown Periods. *J. Xidian Univ.*, 15(8): 628-650 (2021)

Fuhrman, S.: A note on the $M/G/1$queue with server vacations. *Oper. Res.* 32, 1368-1373 (1981)

Gupta, P.K. and Hira, D.S.: *Operations Research*. Revised Edition. S.Chand and Co (2013)

Hagighi, A.M., Medhi, J., Mohanty S.G.: On a multiserver markovian queuing system with balking and reneging. *Computer and Oper. Res.* 13, 425-421 (1986)

Haight, F.A. Queueing with balking. Biometrika 44, 360-369 (1957)

Hillier, F. S. & Lieberman, G. J.: *Introduction to operations research*. 8th edn. New York: McGraw-Hill (2005).

Ibe, O.C.: $M/G/1$ vacation queueing systems with server time. *American J. Oper. Res.* 5, 77-88 (2014)

Kashyap, B.R.K, Chaudhry M.L.: *An introduction to queueing theory*. A and A Publications, Kingston, Ont. Canada (1988)

Keilson, J., Kooharian, A.: Time dependent queueing processes. *Ann. Math. Stat.* B1104-112 (1960)

Keilson, J., Servi, L.D.: Dynamics of the $M/G/1$vacation model. *Oper. Res.* 35(4), 575–582 (1987)

Kumar, R., Sharma, S.K.: Queuing with reneging, balking and retention of reneged customers. *Intl J. Math Mod and Meth in Appl. Sci.* 6(7), 819-828 (2012)

Lee, S.S., Lee, H.W., Yoon, S.H., Chae, K.C.: Batch arrival queue with N-policy and single vacation. *Comput. and Oper. Res.* 22(2), 173-189 (1995)

Levy, Y., Yechilai, U.: Utilization of idle time in an $M/G/1$ queueing system. Manage. Sci. 22(2), 202-211 (1975)

Li, H., Zhu, Y.: Analysis of $M/G/1$queues with delayed vacations and exhaustive service discipline. *European J. Oper. Res.* 92(1), 125-134 (1996)

Little, J.D.C.: A Proof of the Formula $L = \lambda W$. *Oper. Res.* 9:383-387 (1961)

Madan, K.C., Al-Rawi, Z.R., Al-Nasser, A.D.: On $M^X/\begin{pmatrix} G_1 \\ G_1 \end{pmatrix}/1/G(BS)/V_s$ vacation queue with two types of general heterogeneous service. *J. Appl. Math. and Decision Sciences* 3, 123–135 (2005)

Madan, K.C., Choudhury, G.: A single server queue with two phases of heterogeneous service under Bernoulli schedule and a general vacation time. *Inf. Manage. Sci.* 16(2),1-16 (2005)

Madan, K.C.: Balking phenomenon in the $M^X/G/1/$ vacation queue. *J. Korean Stat. Soc.* 31(4), 491-507 (2002)

Mahanta, S., Choudhury, G.: On $M/\begin{pmatrix} G_1 \\ G_1 \end{pmatrix}/1/V(MV)$queue with two types of general heterogeneous service with Bernoulli feedback. *Cogent Math. Stat.* 5(1), 1-9 (2018)

Maragathasundari, S., Srinivasan, S., Ranjitham, A.: A batch arrival non markovian queue with three types of service. *Intl. J. Comp. Appl.* 83(5), 43-47 (2013)

Maragathasundari, S., Srinivasan, S.: Transient analysis of $M/G/1$ queue with Bernoulli feedback and three types of service. *Intel. Conf. Mathl. mod. Appl. Soft. Comp.*, CIT, Coimbatore, India 2 (2012)

Medhi, J.: *Stochastic models in queueing theory*, 2nd ed. Elsevier Sci., USA (2003)

Murthy, P.R.: *Operations Research*, 2nd Edition, New Age International Ltd., Publishers, New Delhi, India (2007)

Rosenberg, E., Yechiali, U.: The $M^X/G/1$ queue with single and multiple vacations under LIFO service regime. *Oper. Res. Lett.* 14(3), 171–179 (1993)

Ross, S.M.: *Introduction to Probability Models*, 2nd ed., New York: Academic Press (1980)

Shanthikumar, J.G.: On stochastic decomposition in the $M/G/1$ type queues with generalized vacations. *Oper. Res.* 36, 566-569 (1988)

Sharma J.K.: *Operations research: theory and applications,* 5th ed., Macmillan India Ltd New Delhi (2013)

Takagi, H.: *Queuing Analysis: A foundation of performance evaluation*. Vacation and priority systems 1(1), North Holland, Amsterdam (1991)

Whitt, W. (1983). Comparing batch delays and customer delays. *Bell. Sys. Tech. J.* 62, 2001-(2009).

Winston, W.L.: *Operations research applications and algorithms*, 4ed., Brooks/Cole-Thomson Learning Nelson 1120 Birchmount Road Toronto, Ontario M1K 5G4 Canada (2004)

Zhang, Z., Vickson, R.G.: A simple approximation for mean waiting time in $M/G/1$ queue with vacations and limited service discipline. *Oper. Res. Lett.* 13(1), 21-26 (1993)