Journal of Advances in Mathematics and Computer Science

xx(x): ....., 20yy; Article no.JAMCS.xxxxx

ISSN: 2456-9968

(Past name: British Journal of Mathematics & Computer Science, Past ISSN: 2231-0851)

Analysis of a two-server queue with consultation in Markovian environment

Original Research Article

Received: XX December 20XX
Accepted: XX December 20XX
Online Ready: XX December 20XX

Abstract

Multi server queueing models with consultation by one well trained and experienced main server to the fellow servers plays an important role in many modern service systems. In this paper, we analyze a two-server queueing model, where the main server provides consultation support in Markovian environment to the regular server. The main server serves customers directly but also extends consultation to the regular server with preemptive priority over its own customer service. The service of the customer at the main server will be interrupted due to higher-priority consultations. Both the number of interruptions at the main server and the number of consultations to the regular server are limited to finite upper bounds. The consultations are arised from M Markovian environmental factors which are related through a transition probability matrix F. The arrival process and consultation requirement are governed by two mutually independent Poisson processes. The service times at the main server and the regular server are modelled as mutually independent phase type distributions. The duration of the threshold clock is exponentially distributed whereas that of the super clock is a phase-type distributed parameter. The stability condition for the system is established and some performance measures are investigated through numerical analysis.

Keywords: main server, regular server, consultation, interruption, Markovian environmental factors

1 Introduction

Chakravarthy (1) introduces the concept of a multi-server queueing system in which timely consultations are offered by one of the main servers to the other servers. This well-experienced server is capable of resolving the doubts of the other servers, and thereby ensuring smoother and more efficient service. Both servers serve the customers from a single queue, but the main server provides pre-emptive priority for consultation over service. In this situation, the service of the customer at the main server is subjected to interruption, and that customer is forced to wait until the consultation process is completed and after completion of consultation, the service of that customer will be continued. The service of the customer at the regular server is not regarded as interrupted, since consultation is incorporated as a component of the service process to maintain the quality of service at the regular server.

Such queueing systems can be observed in hospitals, petrol pumps, banks, and similar service environments. In a hospital setting, the general phisycians (regular servers) examine the patients in a queue. If a patient exhibits any rare or complex symptoms, the physician consults a specialist (main server) for advice. The specialist provides appropriate consultation to the physician, which may interrupt the ongoing service at the specialist. Within a banking system, a teller (regular server) handles routine customer transactions such as deposits, withdrawals, and loan-related services. When a transaction appears unusual, potentially fraudulent, or operationally complex to him, the teller seeks consultation from the branch manager (main server). The branch manager immediately provides guidance to the teller before proceeding the transaction. A customer care representative (regular server) attends to customer inquiries. When a query involves a specific policy or requires specialized knowledge beyond his expertise, the representative consults a specialized officer (main server) for clarification or guidance. Such timely consultations are essential for maintaining a high service quality.

In the paper (2), Klimenock et. al analyse a multi-server queueing model with heterogeneous customers. There are two types of customers with different priorities. the servers follow independent phase type distributions for their service times.

Samouylov et. al. (3) considers a multi-server queueing model that serves two correlated streams of requests. A non-preemptive priority mechanism is implemented by introducing a preliminary delay for one stream in intervening buffers, from which requests are extracted at different rates.

T. Resmi and K Ravikumar (4) analyse a three server queue with consultations offered by a main server to the other two regular servers. There is a finite buffer at the main server, where the main server provides pre-emptive priority for consultation over the customer service.

In (5), Resmi et. al. narrate a two server queueing model with mutual consultations. Here both the servers offer timely consultations to the other while the service times at the servers follow independent phase type distributions.

Krishnamoorthy et. al. (6) investigates a two-server queueing model where the service times at both servers are modeled by phase-type distributions of identical order, although the service rate of one server is lower than that of the other.

In the paper (7), Ayyappan G. and Archana G. elaborates a classical queueing system with two types of heterogeneous servers. Here optional services are offered by the servers to the customers, if they are unsatisfied. The service at the server 2 may get interrupted due to breakdown during any type of service. Even though there is a breakdown, server 2 is capable of continue and finish the service of the current customer at a slower rate.

In Krishnamoorthy et.al (8), we can see a single server queueing model with interruptions to the server. The number of interruptions are controlled by a super clock and a finite upper bound.

White and Christie were the first to study queueing systems with service interruptions in their paper (9). The service of the customer resumes immediately after the interruption ends.
Gaver (10) , Keilson (11), Ibe and Trivedi (12), Avi-Izhak and Naor (13) and Fiems et. al (14) are some papers which discuss queueing systems with generally distributed service and duration of interruptions.

Bhaskar Senguptha (15) discusses a queueing system operating in an alternating random environment. The server faces breakdown randomly and is unable to serve customers during it's repair. A few of the customers arriving during the period of break down are directed to another serving counter.

Krishnamoorthy et.al (16) considers a queueing model with two servers rendering consultations in a random environment.

In the present paper, we consider the consultations due to some Markovian environmental factors where these factors are related by a transition matrix. The number of interruptions to a customer at the main server and the number of consultations provided to the regular server are each restricted by finite upper bounds. The main server immediately offers consultation to the regular server if either the number of interruptions to the customer at the main server has not reached its upper bound and the super clock has not expired. Otherwise the regular server is forced to wait until the main server completes the customer's service. The threshold clock is a mechanism which determines whether the services at both the servers are to be restarted or resumed.

2 Description of model

In this model we consider a queueing system consisting of two servers: one main server and one regular server. The arrival process of the customers to the system is a Poisson process, where its rate is λ. The service times of customers at the main and regular servers are modelled as independent phase type distributions characterised by ($\boldsymbol{\alpha}$,V) and ($\boldsymbol{\beta}$,Y) respectively. These distributions have number of phases a and b, respectively. Denote $V^0 = -V\mathbf{e}$ and $Y^0 = -Y\mathbf{e}$ where **e** represents a column vector all of whose entries are equal to one, with appropriate dimension.
The main server extends consultation services to the regular server whenever required. Let $f_1, f_2, ..., f_M$ be M environmental factors which represent external or system-dependent conditions under which the regular server requires consultation from the main server. The occurrence of a consultation is assumed to depend on the prevailing environmental factor at a given time. The environmental factors evolve according to a Markov chain with a transition probability matrix F. The occurrences of consultation requests form a Poisson process with rate θ, where the probability of the i^{th} factor is δ_i, $i = 1, 2, ..., M$. The consultation time, when the system is under the environmental factor f_i, follows an exponential distribution with rate ξ_i. A threshold clock is introduced to regulate the restart or resumption of the services at both servers. Upon interruption, the service at the main and regular servers is either resumed or restarted depending on the expiry of the threshold time. If the threshold clock has not expired, then the services at both servers are resumed from the point at which they were interrupted; otherwise, the services are restarted afresh at both servers. The threshold time is assumed to follow an exponential distribution with rate ω.

A maximum of K interruptions is permitted for a customer being served at the main server, and a maximum of L consultations is needed for the regular server. The super clock duration is assumed to follow a phase-type distribution characterized by the representation (γ, G) with number of phases c, where $\mathbf{G}^0 = -\mathbf{Ge}$.

Notations :- The following notations are used throughout the model:

- $K_0 = K(c + 1)$ and $K_1 = K_0 + 1$
- $L_1 = LM + 1$, $L_2 = 2LM$

- $D_1 = L_1K_1ab + L_2b + L_2K_0ab + L_2ab$
- $D_0 = K_1a + L_1b + L_2b$
- $\tilde{\boldsymbol{\gamma}} = (\boldsymbol{\gamma}, 0), \tilde{\boldsymbol{\eta}} = (1, 0)$
- $\dot{I}_1 = \begin{bmatrix} \mathbf{0} & I_{K_0} \end{bmatrix}_{K_0 \times K_1}, \dot{I}_2 = \begin{bmatrix} \mathbf{0} & I_{LM} \end{bmatrix}_{LM \times L_1}$
- $\tilde{G} = \begin{bmatrix} G & G^0 \\ \underline{\mathbf{0}} & 0 \end{bmatrix}$ and $F^* = \begin{bmatrix} \boldsymbol{\delta} & \underline{\mathbf{0}} \\ O & I_{L-1} \otimes F \\ O & O \end{bmatrix}$
- $\boldsymbol{\xi} = (\xi_1, \xi_2, ..., \xi_K)^{'}$

The queueing model is $X = \{X(\tau), \tau \geq 0\}$,
where $X(\tau) = \{N(\tau), \sigma(\tau), C_1(\tau), E(\tau), C_2(\tau), V_1(\tau), V_2(\tau), Q_1(\tau), Q_2(\tau)\}$.

The relevant variables used in the model are defined below:

$N(\tau)$	–	the number of customers in the system
$C_1(\tau)$	–	number of consultations already received by the regular server while the service of a given customer
$C_2(\tau)$	–	the number of interruptions that have occurred to a customer currently being served at the main server

$V_1(\tau), Q_1(\tau)$ and $Q_2(\tau)$ represent the phases of the super clock the main server and the regular server, respectively.

Here $\sigma(\tau)$ denotes the status of the servers at time τ such that

$$\sigma(\tau) = \begin{cases} \tilde{0}, & \text{if the regular server is busy and the main server is idle} \\ 0, & \text{if the regular server is either busy or not and the main server is busy} \\ 1, & \text{if the main server is engaged solely in providing consultation} \\ 2, & \text{If the main server is engaged in providing consultation,} \\ & \text{while its own customer is in an interrupted state} \\ 3, & \text{If the regular server is waiting for consultation,} \\ & \text{while the main server completes its current service.} \end{cases}$$

$V_2(\tau)$ represents the status of the threshold clock. It take values 0 and 1 corresponding to the expiry and running , respectively of this clock.

The variable $E(\tau)$ represents the environmental factor associated with consultation. Specifically, if the regular server is serving a customer following a consultation, then $E(\tau)$ denotes the environmental factor that caused that consultation. If the regular server is either receiving consultation or waiting to receive consultation, then $E(\tau)$ denotes the environmental factor governing the ongoing consultation. Consequently, $E(\tau) = j$, where $1 \leq j \leq M$ in the following situations:
(1)$N(\tau) = 1$ and $\sigma(\tau) = \tilde{0}$ or $N(\tau) \geq 2$ and $\sigma(\tau) = 0$ with $1 \leq C_1(\tau) \leq L$
(2)$N(\tau) = 1$ and $\sigma(\tau) = 1$
(3)$N(\tau) \geq 2$ and $\sigma(\tau) = \{1, 2, 3\}$.

The variable $C_1(\tau)$ is 0 signifies that the regular server has not obtained any consultation. Consequently, the environmental factor $E(\tau)$ is not applicable when $C_1(\tau) = 0$ and its phases are not considered in this case.

$\{X(\tau), t \geq 0\}$ is a Continuous Time Markov Chain with state space

$$\zeta = \{0\} \cup \bigcup_{n=1}^{\infty} \zeta(n).$$

The terms $\zeta(n)$'s are defined as
$\zeta(1) = \zeta(1,0) \cup \zeta(1,\tilde{0}) \cup \zeta(1,1)$ and
$\zeta(n) = \zeta(n,0) \cup \zeta(n,1) \cup \zeta(n,2) \cup \zeta(n,3)$, for $n \geq 2$,
where
$\zeta(1,0) = \{(1,0,0,\tau_1)\} \cup \{(1,0,p,t_1,\tau_1) : 1 \leq p \leq K\}$,
$\zeta(1,\tilde{0}) = \{(1,\tilde{0},0,\tau_2) \cup (1,\tilde{0},m,t_3,\tau_2) : 1 \leq m \leq L\}$,
$\zeta(1,1) = \{(1,1,m,t_2,t_3,\tau_2) : 0 \leq m \leq L-1\}$,
$\zeta(n,0) = \{(n,0,0,0,\tau_1,\tau_2) \cup (n,0,m,t_3,0,\tau_1,\tau_2) \cup (n,0,0,p,t_1,\tau_1,\tau_2)$
$\cup\, (n,0,m,t_3,p,t_1,\tau_1,\tau_2) : 1 \leq m \leq L, 1 \leq p \leq K\}$,
$\zeta(n,1) = \{(n,1,m,t_2,t_3,\tau_2) : 0 \leq m \leq L-1\}$,
$\zeta(n,2) = \{(n,2,m,t_2,t_3,p,t_1,\tau_1,\tau_2) : 0 \leq m \leq L-1, 0 \leq p \leq K-1\}$,
$\zeta(n,3) = \{(n,3,m,t_2,t_3,\tau_1,\tau_2) : 0 \leq m \leq L-1\}$,
for $n \geq 2$ with $0 \leq t_1 \leq c$, $t_2 = \{1,0\}$, $1 \leq t_3 \leq M$,
$1 \leq \tau_1 \leq a$ and $1 \leq \tau_2 \leq b$.
The infinitesimal generator P is given by

$$P = \begin{bmatrix} -\lambda & A_1 & & & & \\ A_2 & A_3 & A_4 & & & \\ & A_5 & B_1 & B_0 & & \\ & & B_2 & B_1 & B_0 & \\ & & & \ddots & \ddots & \ddots \\ & & & & & \end{bmatrix} \tag{2.1}$$

where

$$A_1 = \lambda \begin{bmatrix} \mathbf{e}'_{K_1}(1) \otimes \boldsymbol{\alpha} & \underset{\sim}{\mathbf{0}} \end{bmatrix},\ A_2 = \begin{bmatrix} \mathbf{e}_{K_1} \otimes V^0 \\ \mathbf{e}_{LM+1} \otimes Y^0 \\ \mathbf{0} \end{bmatrix},$$

$$A_3 = \begin{bmatrix} I_{K_1} \otimes V & O & O \\ O & A_{31} & \theta F^* \otimes \tilde{\boldsymbol{\eta}} \otimes I_b \\ O & \dot{I}_2 \otimes \xi \otimes \Delta & I_L \otimes \nabla \otimes I_b \end{bmatrix} - \lambda I,$$

$$A_4 = \lambda \begin{bmatrix} A_{41} & A_{42} & O \end{bmatrix},\ A_5 = \begin{bmatrix} A_{51} & A_{52} & A_{53} \end{bmatrix},$$

$$B_0 = \lambda I,\ B_1 = \begin{bmatrix} B_{11} & O & B_{12} & B_{13} \\ B_{14} & B_{15} & O & O \\ B_{16} & O & B_{17} & O \\ O & O & O & B_{18} \end{bmatrix} - \lambda I,$$

$$B_2 = \begin{bmatrix} B_{21} & B_{53} & O \end{bmatrix}.$$

Order of the square matrices B_0, B_1 and B_2 is D_1;
that of A_3 is D_0. The orders of A_1, A_2, A_4, A_5 are
are $1 \times D_0$, $D_0 \times 1$, $D_0 \times D_1$ and $D_1 \times D_0$, respectively.
The block matrices are described as follows:
$A_{31} = I_{L_1} \otimes Y - \theta I^* \otimes I_b$,

$$A_{41} = \begin{bmatrix} \mathbf{e}'_{L_1}(1) \otimes I_{K_1} \otimes I_a \otimes \boldsymbol{\beta} \\ I_{L_1} \otimes \mathbf{e}'_{K_1}(1) \otimes \boldsymbol{\alpha} \otimes I_b \\ O \end{bmatrix}_{C_0 \otimes K_1 M_1 ab},\ A_{42} = \begin{bmatrix} O \\ I_{L_2 b} \end{bmatrix}_{D_0 \otimes L_2 b},$$

$$A_{51} = \begin{bmatrix} \mathbf{e}_{L_1} \otimes I_{K_1} \otimes I_a \otimes Y^0 \\ O \end{bmatrix}_{D_1 \otimes K_1 a},$$

$$A_{52} = \begin{bmatrix} I_{L_1} \otimes \mathbf{e}_{K_1} \otimes V^0 \otimes I_b \\ O \end{bmatrix}_{D_1 \otimes L_1 b},$$

$$A_{53} = \begin{bmatrix} O \\ I_{L_2} \otimes V^0 \otimes I_b \end{bmatrix}_{D_1 \otimes L_2 b},$$

$B_{11} = I_{L_1} \otimes I_{K_1} \otimes (V \oplus Y) - \theta I^* \otimes I_{K_1} \otimes I_{ab},$

$B_{12} = \theta F^* \otimes J \otimes \tilde{\boldsymbol{\eta}} \otimes I_{ab},\ B_{13} = \theta F^* \otimes J^* \otimes \tilde{\boldsymbol{\eta}} \otimes I_{ab},$

$B_{14} = \dot{I}_2 \otimes \xi \otimes \tilde{\Delta},\ B_{15} = I_L \otimes \nabla \otimes I_b,\ B_{16} = \dot{I}_2 \otimes \xi \otimes \dot{I}_1 \otimes \Delta^*,$

$B_{17} = I_L \otimes \tilde{\nabla} \otimes I_{ab},\ B_{18} = I_{LM} \otimes (H \oplus V) \otimes I_b,$

$B_{21} = \begin{bmatrix} \tilde{V}^0 + \tilde{Y}^0 \\ O \end{bmatrix}_{D_1 \otimes L_1 K_1 ab}.$

Here

$$I^* = \begin{bmatrix} I_{(L-1)M+1} & O \\ O & O \end{bmatrix}_{L_1 \otimes L_1},$$

$$J = \begin{bmatrix} diag(\tilde{\boldsymbol{\gamma}}, I_{K-1} \otimes \hat{I}_c) \\ O \end{bmatrix},\ J^* = \begin{bmatrix} 0 \\ \mathbf{e}_{K-1} \otimes \hat{\mathbf{e}}_c \\ \mathbf{e}_{c+1} \end{bmatrix},$$

$$\tilde{V}^0 = I_{L_1} \otimes \mathbf{e}_{K_1} \otimes V^0 \otimes \boldsymbol{\alpha} \otimes I_b,\ \tilde{Y}^0 = \mathbf{e}_{LM+1} \otimes I_{K_1} \otimes I_a \otimes Y^0 \otimes \boldsymbol{\beta},$$

$$\Delta = \begin{bmatrix} I_b \\ \mathbf{e}_b \otimes \boldsymbol{\beta} \end{bmatrix},\ H = \omega \begin{bmatrix} -1 & 1 \\ 0 & 0 \end{bmatrix},\ \nabla = I_L \otimes (\boldsymbol{\xi} \oplus H) \otimes I_b,$$

$$\tilde{\nabla} = I_L \otimes [\boldsymbol{\xi} \otimes I_{2K_0} + I_{MK} \otimes (\tilde{G} \oplus H)] \otimes I_{ab},$$

$$\tilde{\Delta} = \begin{bmatrix} \mathbf{e}'_{K_1}(1) \otimes \boldsymbol{\alpha} \otimes I_b \\ \mathbf{e}_b \otimes \mathbf{e}'_{K_1}(1) \otimes \boldsymbol{\alpha} \otimes \boldsymbol{\beta} \end{bmatrix},\ \Delta^* = \begin{bmatrix} I_{ab} \\ \mathbf{e}_{ab} \otimes \boldsymbol{\alpha} \otimes \boldsymbol{\beta} \end{bmatrix}.$$

3 Steady state analysis

In this section, we carry out the steady-state analysis of the queueing model under consideration. We begin by deriving the stability condition for the system.

3.1 Stability condition

Let $\boldsymbol{\pi}$ denotes the steady-state probability vector of the generator matrix $B_0+B_1+B_2$ of the underlying Markov process.
The LIQBD structure of the model implies that the queueing system is stable, (see, Neuts (17)) if and only if

$$\boldsymbol{\pi} B_0 \mathbf{e} < \boldsymbol{\pi} B_2 \mathbf{e}. \tag{3.1}$$

The steady-state probability vector $\boldsymbol{\pi}$ does not admit a closed-form expression in terms of the system parameters.

Define the traffic intensity ρ as

$$\rho = \frac{\boldsymbol{\pi} B_0 \mathbf{e}}{\boldsymbol{\pi} B_2 \mathbf{e}}. \tag{3.2}$$

It is noted that the stability condition in equation (3.1) reduces to the familiar condition $\rho < 1$ where denotes the traffic intensity of the system. We will examine the influence of the input parameters of the model on the traffic intensity of the system in Section 4.

3.2 Steady state probability vector

Since the model is formulated as a quasi-birth–death (QBD) process, its steady-state distribution possesses a matrix-geometric form, provided that the stability condition is satisfied. We assume that the stability condition (3.1) is satisfied. Let us denote the steady-state probability vector of the generator P given in equation (2.1) by $\boldsymbol{p}$.

Partitioning $\boldsymbol{p}$ as

$$\boldsymbol{p} = (\boldsymbol{p}_0, \boldsymbol{p}_1, \boldsymbol{p}_2, ...) \tag{3.3}$$

we see that the sub-vectors of $\boldsymbol{p}$, under the assumption that the stability condition (3.1) holds, are obtained as (see, Neuts (17))

$$\boldsymbol{p}_n = \boldsymbol{p}_2 R^{n-2}, n \geq 3 \tag{3.4}$$

where R is the minimal non-negative solution to the matrix quadratic equation:

$$R^2 B_2 + R B_1 + B_0 = 0. \tag{3.5}$$

$\boldsymbol{p}_0$, $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$ are obtained using the boundary equations

$$\begin{aligned} -\lambda \boldsymbol{p}_0 + \boldsymbol{p}_1 A_2 &= 0 \\ \boldsymbol{p}_0 A_1 + \boldsymbol{p}_1 A_3 + \boldsymbol{p}_2 A_5 &= 0 \\ \boldsymbol{p}_1 A_4 + \boldsymbol{p}_2 (B_1 + R B_2) &= 0 \end{aligned} \tag{3.6}$$

From the normalizing condition we get

$$\boldsymbol{p}_0 + \boldsymbol{p}_1 \mathbf{e} + \boldsymbol{p}_2 (I - R)^{-1} \mathbf{e} = 1. \tag{3.7}$$

Once we obtain the rate matrix R, the steady-state probability vector $\boldsymbol{p}$ can be evaluated by utilizing the particular structure of the underlying coefficient matrices.

3.3 Performance measures

In this section, we present a set of important system performance measures to illustrate the qualitative behavior of the model under study. The relevant measures are listed below together with their analytical expressions for computation.

Towards this end, we further partition the vectors $\boldsymbol{p}_n$ as $\boldsymbol{p}_1 = (\boldsymbol{p}_{10}, \boldsymbol{p}_{1\tilde{0}}, \boldsymbol{p}_{11})$ and $\boldsymbol{p}_n = (\boldsymbol{p}_{n0}, \boldsymbol{p}_{n1}, \boldsymbol{p}_{n2}, \boldsymbol{p}_{n3})$, where $n \geq 2$.

Note that $\boldsymbol{p}_0$ is a scalar, $\boldsymbol{p}_1 = (\boldsymbol{p}_{10}, \boldsymbol{p}_{1\tilde{0}}, \boldsymbol{p}_{11})$ and $\boldsymbol{p}_n = (\boldsymbol{p}_{n0}, \boldsymbol{p}_{n1}, \boldsymbol{p}_{n2}, \boldsymbol{p}_{n3})$, where $n \geq 2$.
Here $\boldsymbol{p}_{10}, \boldsymbol{p}_{1\tilde{0}}, \boldsymbol{p}_{11}, \boldsymbol{p}_{n0}, \boldsymbol{p}_{n1}, \boldsymbol{p}_{n2}, \boldsymbol{p}_{n3}$, for $n \geq 2$ are vectors of dimensions $M_1 a$, $K_1 b$, $K_2 b$, $K_1 M_1 ab$, $K_2 b$, $M_0 K_2 ab$ and $K_2 ab$, respectively.

We now proceed to evaluate the performance measures to analyze the behavior of the system.

(1) Mean number of customers in the system

$$E_1 = \sum_{n=1}^{\infty} n \boldsymbol{p}_n \mathbf{e}. \tag{3.8}$$

(2) Mean number of customers in the queue

$$E_2 = \sum_{n=2}^{\infty} (n-1) \boldsymbol{p}_{n1} \mathbf{e} + \sum_{n=3}^{\infty} (n-2)(\boldsymbol{p}_{n0} \mathbf{e} + \boldsymbol{p}_{n2} \mathbf{e} + \boldsymbol{p}_{n3} \mathbf{e}). \tag{3.9}$$

(3) Effective consultation rate

$$E_3 = \theta \sum_{m=0}^{K-1} \boldsymbol{p}_{1\tilde{0}m} \mathbf{e} + \theta \sum_{n=2}^{\infty} \sum_{m=0}^{K-1} \boldsymbol{p}_{n0m} \mathbf{e}. \tag{3.10}$$

(4) Effective interruption rate

$$E_4 = \theta \sum_{n=2}^{\infty} \sum_{m=0}^{K-1} \boldsymbol{p}_{n0m0}\mathbf{e} + \theta \sum_{n=2}^{\infty} \sum_{m=0}^{K-1} \sum_{k=1}^{M-1} \sum_{l_1=1}^{c} \boldsymbol{p}_{n0mkl_1}\mathbf{e} \tag{3.11}$$

(5) Probability that the main server is idle

$$\Gamma_{mi} = \boldsymbol{p}_0\mathbf{e} + \boldsymbol{p}_{1\tilde{0}}\mathbf{e}. \tag{3.12}$$

(6) Probability that the regular server is idle

$$\Gamma_{ri} = \boldsymbol{p}_0\mathbf{e} + \boldsymbol{p}_{10}\mathbf{e}. \tag{3.13}$$

(7) Probability that the main server is busy serving a customer

$$\Gamma_{mb} = \boldsymbol{p}_{10}\mathbf{e} + \sum_{n=2}^{\infty} \boldsymbol{p}_{n0}\mathbf{e} + \sum_{n=2}^{\infty} \boldsymbol{p}_{n3}\mathbf{e}. \tag{3.14}$$

(8) Probability that the regular server is busy serving a customer

$$\Gamma_{rb} = \boldsymbol{p}_{1\tilde{0}}\mathbf{e} + \sum_{n=2}^{\infty} \boldsymbol{p}_{n0}\mathbf{e}. \tag{3.15}$$

(9) Probability that the regular server is receiving consultation

$$\Gamma_{rc} = \sum_{n=1}^{\infty} \boldsymbol{p}_{n1}\mathbf{e} + \sum_{n=2}^{\infty} \boldsymbol{p}_{n2}\mathbf{e}. \tag{3.16}$$

(10) Probability that the regular server is waiting to receive consultation

$$\Gamma_{wc} = \sum_{n=2}^{\infty} \boldsymbol{p}_{n3}\mathbf{e}. \tag{3.17}$$

(11) Probability that the main server stays interrupted

$$\Gamma_{min} = \sum_{n=2}^{\infty} \boldsymbol{p}_{n2}\mathbf{e}. \tag{3.18}$$

4 Numerical examples

Let $\boldsymbol{\alpha} = \begin{bmatrix} 0.3 & 0.7 \end{bmatrix}$; $Y = \begin{bmatrix} -12 & 6 \\ 5 & -10 \end{bmatrix}$; $\boldsymbol{\beta} = \begin{bmatrix} 0.4 & 0.6 \end{bmatrix}$; $V = \begin{bmatrix} -9 & 3 \\ 2 & -8 \end{bmatrix}$;
$\boldsymbol{\gamma} = \begin{bmatrix} 0.6 & 0.4 \end{bmatrix}$; $G = \begin{bmatrix} -12 & 8 \\ 8 & -12 \end{bmatrix}$; $\boldsymbol{\delta} = \begin{bmatrix} 0.3 & 0.5 & 0.2 \end{bmatrix}$;
$L = 3$; $K = 3$; $\omega = 2$; $\boldsymbol{\xi} = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}^T$; $F = \begin{bmatrix} 0.3 & 0.3 & 0.4 \\ 0.4 & 0.5 & 0.1 \\ 0.3 & 0.4 & 0.3 \end{bmatrix}$.

It is noted that, for the above vectors, matrices and parameter values, the stability condition $\rho < 1$ is satisfied.

From Table 1 it is observed that as the consultation rate θ increases, then traffic intensity ρ will increase and consequently E_3 and E_4 will increase. Thus, there is an increase in Γ_{min} and Γ_{rc}. As increases, the frequency of consultations rises, which accelerates the attainment of the maximum

Table 1: Effect of θ on the various performance measures

$\lambda = 3$

θ	3	3.5	4	4.5	5
ρ	0.7211	0.7849	0.8445	0.9001	0.9519
E_1	4.4827	5.9070	7.6951	9.9176	12.6401
E_2	3.3316	4.5717	6.0350	7.6359	9.2266
E_3	0.3886	0.4544	0.5156	0.5706	0.6186
E_4	0.5579	0.6477	0.7288	0.7971	0.8482
Γ_{mi}	0.3388	0.2896	0.2446	0.2045	0.1696
Γ_{ri}	0.4435	0.3799	0.3214	0.2692	0.2236
Γ_{mb}	0.3135	0.3053	0.2962	0.2859	0.2737
Γ_{rb}	0.1947	0.1969	0.1972	0.1981	0.1993
Γ_{min}	0.2461	0.2882	0.3263	0.3580	0.3811
Γ_{rc}	0.3471	0.4030	0.4532	0.4950	0.5256
Γ_{rw}	0.0140	0.0181	0.0222	0.0261	0.0294

allowable interruptions at the main server or triggers the realization of the super clock at an earlier stage. As a result, the main server is forced to prioritize the completion of its ongoing service before engaging in additional consultations. This leads to an increase in the waiting time experienced by the regular server for receiving consultation. Thus Γ_{rw} increases.Furthermore, the likelihood of service restart at the regular server also increases.and hence Γ_{rb} increases. An increase in Γ_{min}, Γ_{rc} and Γ_{rw} leads to an increase in the waiting time, causing customers to remain in the system and the queue for extended periods. Consequently, this results in an increase in E_1 and E_2. So the idle times Γ_{mi} and Γ_{ri} of the main server and the regular server, respectively decrease. Also note that, as the main server is required to devote more time to consultations, thereby reducing the time available for serving customers.Therefore Γ_{mb} decreases.

Table 2: Effect of λ on the various performance measures

$\theta = 2$

λ	3	3.5	4	4.5	5
ρ	0.5808	0.6775	0.7743	0.8711	0.9679
E_1	2.4768	3.9794	6.2527	9.6005	14.3406
E_2	1.5690	2.8554	4.7760	7.2039	9.6201
E_3	0.2504	0.3313	0.4116	0.4848	0.5465
E_4	0.3657	0.4607	0.5479	0.6166	0.6544
Γ_{mi}	0.4446	0.3535	0.2716	0.2024	0.1471
Γ_{ri}	0.5796	0.4695	0.3662	0.2763	0.2030
Γ_{mb}	0.3282	0.3599	0.3853	0.4015	0.4036
Γ_{rb}	0.1865	0.2349	0.2794	0.3143	0.3334
Γ_{min}	0.1573	0.2079	0.2569	0.2969	0.3199
Γ_{rc}	0.2272	0.2861	0.3397	0.3805	0.4004
Γ_{rw}	0.0067	0.0090	0.0113	0.0131	0.0142

Table 2 shows that as the arrival rate increases, the traffic intensity increases accordingly. This results in a higher number of customers in the system and, consequently, greater accumulation and congestion in the queue. So E_1 and E_2 increase. As a result, E_3 and E_4 will also increase. Thus there is a hike in Γ_{min} and Γ_{rc}. Hence Γ_{rw} also increases. Since the arrival rate is increased and the queue is equipped with more customers, the service time of the servers increases. Thus Γ_{mb} and Γ_{rb} show an increase. This results in a decrease in Γ_{mi} and Γ_{ri}.

Conclusion

In this paper we analyse a queueing model with consultations in Markovian environment. The system comprises a main server and a regular server, wherein the main server serves customers while also providing consultation to the regular server with preemptive priority over its own customer service. The number of interruptions at the main server is regulated by an upper bound and a super clock mechanism. Consultations are triggered by environmental factors governed by a Markov process with a specified transition probability matrix. The stability condition of the system is derived, and various performance measures are evaluated numerically to examine the system's behavior.

References

[1] Chakravarthy S. R. (2014). A multi-server queueing model with server consultations, Eur.J. of Oper. Res., 233(3), 625-639.

[2] Klimenok, V., Dudin, A., Vishnevsky, V. (2020). Priority multi-server queueing system with heterogeneous customers. Mathematics , 8, 1501.

[3] Samouylov, K., Dudina, O., Dudin, A. (2023). Analysis of Multi-Server Queueing System with Flexible Priorities. Mathematics, 11, 1040.

[4] Thekkiniyedath Resmi., K. Ravikumar. (2021). Three-Server Queue with Consultations by Main Server with a Buffer at the Main Server, Information Technologies and Mathematical Modelling, Queueing Theory and Applications. ITMM 2020. Communications in Computer and Information Science, 1391, 131–142.

[5] Resmi, T., Lakshmy, B., Krishnamoorthy, A. (2018). A Two-Server Queue with Mutual Consultations. J Indian Soc Probab Stat 19, 201–215.

[6] Krishnamoorthy, A., Divya, V. (2020). A Two-Server Queueing System with Processing of Service Items by a Server. In Applied Probability and Stochastic Processes; Springer Nature: Singapore, 307–333.

[7] Ayyappan, G., Archana, G. (2023). Analysis of MAP/PH1, PH2/2 Queueing Model with Working Breakdown, Repairs, Optional Service, and Balking. Appl. Appl. Math. Int. J. (AAM), 18, 1.

[8] Krishnamoorthy A., Pramod P. K., Chakravarthy S. R. (2013). A Note on characterizing service interruptions with phase type distribution, Stochastic Analysis and Applications , 31(4), 671-683.

[9] White H., Christie L. S. (1958). Queuing with Preemptive Priorities or with Breakdown, Operations Research, 6, 79–95.

[10] Gaver D. P., Jacobs P. A., Latouche G. (1984). Finite birth and death models in randomly changing environments, Advances in Applied Probability, 16(4), 715-731.

[11] Keilson J. (1962) . Queues subject to service interruptions, The Annals of Mathematical Statistics 33(4), 1314-1322.

[12] Ibe O. C., Trivedi K. S. (1960). Two queues with alternating service and server breakdown, Queueing Systems 7(3), 253-268.

[13] Avi-Itzhak B, Naor P. (1963). Some queueing problems with the service station subject to breakdowns, Oper. Res. 11(3), 303-320.

[14] Fiems D., Maertens T., Bruneel H. (2008). Queueing systems with different types of interruptions, Eur J Oper Res 188(3), 838-845.

[15] Bhaskar Sengupta. (1990). A queue with service interruptions in an alternating random environment, Oper. Res., 38(2), 308-318.

[16] Krishnamoorthy A., Resmi T., Lakshmy B. (2019). On a Two-Server Queue with Consultation in Random Environment, Information Technologies and Mathematical Modelling., Queueing Theory and Applications. Communications in Computer and Information Science, 1109, 217-229.

[17] Neuts M.F. (1981). Matrix-geometric solutions in stochastic models, An Algorithmic Approach, The Johns Hopkins University Press, Baltimore.
