Neural, Parallel, and Scientific Computations 23 (2015) 459-486

ON PRIORITY QUEUES GENERATED THROUGH CUSTOMER INDUCED SERVICE INTERRUPTION

A. KRISHNAMOORTHY AND MANJUNATH A. S.

Department of Mathematics, Cochin University of Science and Technology,
Cochin-682022, India.
achyuthacusat@gmail.com
Department of Mathematics, Government College Kottayam-686013, India.
manjunadem@gmail.com

ABSTRACT. In this paper we analyze a two priority queueing system generated as under: arrival of customers to high priority($\mathcal{P}_1$) queue constitutes a Poisson process of rate λ. The waiting room has infinite capacity. They are served one at a time according to FIFO discipline. The service time for each customer of this category follows exponential distribution with parameter μ_1. While in service customers have a tendency to interrupt own service. This occurs according to a Poisson process of rate θ_1. Self interrupted customers are sent to an infinite capacity low priority($\mathcal{P}_2$) queue. When at a service completion epoch of a high priority customer, if there is none left behind in $\mathcal{P}_1$ line, then the server goes to serve customers in $\mathcal{P}_2$. Their service time duration has exponential distribution with parameter μ_2. For the two priority system we assume that $\mathcal{P}_2$ customers are not allowed to interrupt their service. Thus the system consists of a high priority waiting line and a second waiting line which is generated from the first. No customers from outside joins $\mathcal{P}_2$. We consider both preemptive and non preemptive service discipline. The joint system state distribution is obtained from which the marginals are computed. Waiting time distribution of both type of customers are derived. We then extend the results to three priority non preemptive case. Finally the case of $N+1$ priorities is briefly discussed.

AMS (MOS) Subject Classification. 60K25.

1. Introduction

In queueing literature priority queues stands for customers belonging to different classes joining distinct waiting lines(one for each class) to receive service. The highest class of customers have priority(preemptive or non preemptive) over the rest; the next in the order gets priority over all lower class customers and so on. The arrival streams are independent. The dependence between queues is only through preference of service. In contrast, the priority queueing system considered in this paper is a highly dependent one even in its evolution. Each waiting line is generated by the customers in the immediately preceding queue except the highest priority customers. The generation of new queues is due to self interruption: the customer in service may

Received October 1, 2015

interrupt his own and consequently sent to next lower priority queue. Self interruption of service by customers ('customer induced interruption' as is coined by Jacob et al. [13]) is discussed in Dudin et al. [2] and Krishnamoorthy, Jacob[6]. Priority queues were first considered by White and Christie [14] as a queue with interruption of service of low priority customers to provide service to higher priority customers. It may be preemptive or non preemptive. A priority queue with preemptive service can be regarded as a queue with service interruption. Jaiswal [3] is on preemptive priority queue with resumption of service of the low priority customer. Time dependent solution in priority queues is discussed in Jaiswal [4]. A detailed discussion of development in priority queues until 1968 is given in Jaiswal [5]. More recent developments on priority queues could be found in Takagi [11].

Cobham [1] considered the non-preemptive priority queue and derived equilibrium expected waiting time. The first published results for the preemptive discipline were by White and Christie [14]. The notion of preemptive distance is introduced by Takagi and Kodera [12] and they analyze preemptive loss priority queues in which customers of each priority class arrive in a Poisson process and have general service time distribution. Customer induced interruption is introduced by Jacob et al.[13] for the single server case, where service interrupted customers are given priority over primary customers; self interrupted customer takes an exponentially distributed time to get out of interruption. This is extended to the multi-server case in Krishnamoorthy and Jacob [6]. All underlying distributions (inter-arrival time, service time, inter-interruption time, interruption fixation time) are assumed to be independent exponential random variables. Dudin et al.[2] extend the above case to Markovian Arrival Process and Phase type service with c servers and negative customers with a few protected service phases.

The priority queue considered by Miller [7] has two waiting lines, each of infinite capacity and served by a single server. The arrival process to the two queues form two independent Poisson streams with parameters λ_1 and λ_2. The low as well as high priority customers, whether in service or in queue, is counted as the number of such customers in the system. The service time duration for high(low) priority customer has exponential distribution with parameter $\mu_1(\mu_2)$. Both preemptive and non-preemptive service disciplines are considered. The system is analyzed as a three dimensional continuous time Markov chain. The condition for system stability is given by $\frac{\lambda_1}{\mu 1} + \frac{\lambda_2}{\mu_2} < 1$. As an extension of the above, Sapna and Stanford [10] studied a single server queue with arrivals from N classes of customers on a non-preemptive priority basis. Each of these arrivals follow independent Poisson processes with rate $\lambda_i, i = 1, 2, ..., N$ and service is class dependent phase type. The capacity of each waiting line is assumed to be infinite. They analyze the queue length and waiting time processes by deriving a matrix geometric solution for the stationary distribution of the underlying Markov chain.

All the above models consider distinct streams of independent Poisson arrivals to the system. In contrast the present paper considers mainly a 2 priority queueing system where input streams are dependent. The high priority($\mathcal{P}_1$) line has input from outside the system (external arrival) according to a Poisson process of rate λ, whereas the low priority($\mathcal{P}_2$) line has input from the high priority waiting line. Thus, low priority queue is generated from within the system. Hence the system that we consider is a highly dependent one as far as the formation of the low priority waiting line is concerned unlike the priority queues with infinite waiting lines that are so far considered in the literature. The same server serves different customers one at a time according to their priority. As an example consider the queue of patients($\mathcal{P}_1$) waiting to consult a physician. A patient while being examined may have to be referred to a specialist. After consulting the specialist the patient returns to the first physician and waits in the second queue($\mathcal{P}_2$). Unlike in [2, 6, 13] here we do not associate any specific distribution for the duration of interruption of a customer; rather we assume that, once an interrupted customer comes to $\mathcal{P}_2$, he is ready to receive service.

We do an extensive analysis in the two priority case: high priority of external (primary or $\mathcal{P}_1$) customers and a second queue (low priority or $\mathcal{P}_2$) of customers who interrupted their service while being served in the high priority queue. With a maximum of a single interruption permitted, we analyze the system as a three dimensional continuous time Markov Chain. Customers from each waiting class will be taken for service according to the head of the queue discipline. When no high priority customer is available at a service completion epoch the server starts service of the head of the low priority queue. By a suitable arrangement and adjustment, we produce an upper triangular (infinite dimensional) rate matrix R. Once this is achieved, we will be in a position to compute the steady state probability vector. Then this is utilized in the computation of performance of the system. The performance measures here, unlike in other set up, will be of a bit of curiosity as well. This is due to the dependence of the second queue on the first for its generation. Having done these, we proceed to the case of 3 queues (one primary and the other two generated from previous higher priority). Finally we briefly extend our results to the case of $N+1$ queues $N \geq 3$. In all these the systems are studied under steady state. Therefore first we establish the condition for stability of the system and then proceed to the analysis. A special feature of the present model, unlike in classical priority models, is that when the server is in P_i queue all P_j queues except $\mathcal{P}_1$ queue turn out to be empty for $i > j$.

The rest of the paper is arranged as under: In section 2, the case of two priorities is extensively analyzed for the preemptive case. Section 3 is devoted to the study of two priority, non preemptive service discipline. The discussions in section 3 is extended to three priority set up in section 4 and finally section 5 provides a brief description of $N+1$ priority system with $N \geq 3$.

2. Two priority -Preemptive case

Model Description: This section considers a single server infinite capacity queuing system in which customers from outside arrive according to a Poisson process with rate λ. Service times of the external customers ($\mathcal{P}_1$) are exponentially distributed with parameter μ_1. Customers in primary queue interrupt their service according to an exponentially distributed time with parameter θ_1, in which case they have to go to the lower priority($\mathcal{P}_2$) queue. Else, complete service and leaves the system forever. Suppose at the time when a $\mathcal{P}_1$ customer leaves the server by self interruption, and hence joins $\mathcal{P}_2$, finds that none is ahead of him and there was none left behind him in $\mathcal{P}_1$. In this case we assume that this customer is not permitted to have an interruption time, rather he is immediately taken for service in $\mathcal{P}_2$. Lower priority customers are taken for service one at a time from the head of the line whenever the queue of external customers is found to be empty at a service completion epoch. The service of such customers is according to a preemptive service discipline following an exponential distribution with parameter μ_2. That is the arrival of a $\mathcal{P}_1$ customer interrupts the ongoing service of a $\mathcal{P}_2$ customer and hence he joins back as the head of the $\mathcal{P}_2$ queue. Consider the case where not more than one interruption is permitted, that is $N = 1$. Let $N_1(t)$ be the number of $\mathcal{P}_1$ customers including the one in service if any, $N_2(t)$ the number of $\mathcal{P}_2$ customers waiting to get service. Whenever $\mathcal{P}_1$ is nonempty, the head of that line will be under service.

Then $\Omega = \{(N_1(t), N_2(t))\,/t \geq 0\}$ is a CTMC with state space $\{(i,j)/i \geq 0, j \geq 0\}$ $\cup \{0\}$. Here 0 represents the state where there is no customer in the system(neither $\mathcal{P}_1$ nor $\mathcal{P}_2$) and $(0,0)$ is the state where a $\mathcal{P}_2$ customer is in service .

The infinitesimal generator Q has as entries block matrices of infinite dimension since the phases (capacity of waiting line for interrupted customers) is infinite. It is given by

$$Q = \begin{pmatrix} B_{00} & B_{01} & & & \\ B_{10} & B_1 & B_0 & & \\ & B_2 & B_1 & B_0 & \\ & & \ddots & \ddots & \ddots \end{pmatrix} \tag{2.1}$$

with

$$B_{00} = \begin{pmatrix} -\lambda & & & & \\ \mu_2 & -(\lambda+\mu_2) & & & \\ & \mu_2 & -(\lambda+\mu_2) & & \\ & & \ddots & \ddots & \\ & & & & \end{pmatrix}, B_{10} = B_2 = \begin{pmatrix} \mu_1 & \theta_1 & & & \\ & \mu_1 & \theta_1 & & \\ & & \mu_1 & \theta_1 & \\ & & & \ddots & \ddots \\ & & & & \ddots \end{pmatrix}$$

$B_{01} = B_0 = \lambda I_\infty$and $B_1 = -(\lambda + \mu_1 + \theta_1) I_\infty$.

We now establish the system stability requirement.

Theorem 2.1. *The condition for stability of the system is* $\rho = \frac{\lambda}{(\mu_1+\theta_1)} + \frac{\lambda\,\theta_1}{(\mu_1+\theta_1)\mu_2} < 1.$

Proof. By interchanging the level and phase in the model, the matrices B_0, B_1 and B_2 are $B_0 = \begin{cases} \theta_1,\ i = 1,2,3...; j = i-1 \\ 0,\ \text{elsewhere} \end{cases}$, $B_1 = \begin{cases} -(\lambda+\mu_2),\ i = j = 0 \\ -(\lambda+\mu_1+\theta_1),\ i = j = 1,2,... \\ \lambda,\ i = 0,1,2,...; j = i+1 \\ \mu_1, i = 1,2,3...; j = i-1 \\ 0,\ \text{elsewhere} \end{cases}$

and $B_2 = \begin{cases} \mu_2,\ i = j = 0 \\ 0,\ \text{elsewhere} \end{cases}$.

Let $\boldsymbol{\pi} = (\pi_0, \pi_1, \pi_2, ...)$ be the steady state probability vector of the matrix $B (= B_0+B_1+B_2)$. Solving the relations $\boldsymbol{\pi} B = 0$ and $\boldsymbol{\pi}\mathbf{e} = 1$, we get $\pi_j = \left(\frac{\lambda}{\mu_1+\theta_1}\right)^j \pi_0, j \geq 1$. As we have a level independent QBD model, the system is stable if $\boldsymbol{\pi} A_0 \mathbf{e} < \boldsymbol{\pi} A_2 \mathbf{e}$, which simplifies to $\rho < 1$. □

The infinitesimal generator Q constitutes a quasi birth and death(QBD) process with exceptional boundary behavior and an infinite number of sub-levels. The matrix geometric form of the steady state distributions for both preemptive and non-preemptive priority single server queues were investigated by Neuts [8] in the case when number of phases in each level is finite. This is extended to blocks of infinite size in Miller [7] and is contained in the following theorem.

Theorem 2.2. *Let* $\boldsymbol{y} = (\boldsymbol{y}_0, \boldsymbol{y}_1, \boldsymbol{y}_2, \ldots)$ *denote the invariant probability vector for the QBD process* Q*, where* $\boldsymbol{y}_i$ *is the probability vector of infinite dimension corresponding to level* i *. Then the solution for* $\boldsymbol{y}$ *possesses a matrix geometric structure*

$$\boldsymbol{y}_{i+1} = \boldsymbol{y}_i R,\ i \geq 1. \tag{2.2}$$

where the rate matrix R *is the minimal non negative solution to*

$$R^2 B_2 + R B_1 + B_0 = 0. \tag{2.3}$$

The matrix geometric structure in equation (2.2) extended to level '0' is

$$\boldsymbol{y}_1 = \boldsymbol{y}_0 \left(\frac{1}{\lambda} B_{01}\right) R. \tag{2.4}$$

Proof. The relations (2.2) and (2.3) are proved in [7].
From $\boldsymbol{y}Q = 0$, the two boundary equations involving $\boldsymbol{y}_0$ are

$$\boldsymbol{y}_0 B_{00} + \boldsymbol{y}_1 B_{10} = 0, \tag{2.5}$$

$$\boldsymbol{y}_0 B_{01} + \boldsymbol{y}_1 [B_1 + RB_2] = 0. \tag{2.6}$$

From (2.3) it follows that

$$R[RB_2 + B_1] = -B_0.$$

Since $B_0 = \lambda I_\infty$, the matrix R is invertible and (2.6) now simplifies to (2.4). □

Theorem 2.3. *The infinite matrix R possesses the Toeplitz structure*

$$R = \begin{pmatrix} r_0 & r_1 & r_2 & r_3 & \cdots \\ 0 & r_0 & r_1 & r_2 & \cdots \\ 0 & 0 & r_0 & r_1 & \cdots \\ 0 & 0 & 0 & r_0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}$$

where r_k are computed as

$$r_0 = \frac{(\lambda + \mu_1 + \theta_1) - \sqrt{(\lambda + \mu_1 + \theta_1)^2 - 4\lambda\mu_1}}{2\mu_1},$$

$$r_1 = \frac{r_0^2 \theta_1}{\sqrt{(\lambda + \mu_1 + \theta_1)^2 - 4\lambda\mu_1}},$$

$$r_k = \frac{\theta_1 \left[\sum_{i=0}^{k-1} r_i r_{k-1-i}\right] + \mu_1 \left[\sum_{i=1}^{k-1} r_i r_{k-i}\right]}{\sqrt{(\lambda + \mu_1 + \theta_1)^2 - 4\lambda\mu_1}}, k > 1.$$

Proof. The structure of the process revealed by matrices in Q^* and the interpretation of rate matrix imply the special structure of R. On expanding (2.3), the following relations are obtained;

$$r_0^2 \mu_1 - (\lambda + \mu_1 + \theta_1) r_0 + \lambda = 0. \tag{2.7}$$

$$\left(\sum_{i=0}^{k-1} r_i r_{k-1-i}\right) \theta_1 + \left(\sum_{i=0}^{k} r_i r_{k-i}\right) \mu_1 - (\lambda + \mu_1 + \theta_1) r_k = 0, \quad k \geq 1. \tag{2.8}$$

Solving these, the expressions for $r_k, k = 0, 1, 2...$, are established. □

2.1. **The Joint and Marginal Probabilities.**

2.1.1. *The Joint Probabilities:* The steady state probability vector $\boldsymbol{y} = (\boldsymbol{y}_0, \boldsymbol{y}_1, \boldsymbol{y}_2, \ldots)$ of Q is computed first. Here $\boldsymbol{y}_0 = (y_0, y_{00}, y_{01}, y_{02}, ...)$ where y_0 is the probability of the idle state and y_{00} is the probability of providing service to a $\mathcal{P}_2$ customer when none is waiting in either queues. $\boldsymbol{y}_i = (y_{i0}, y_{i1}, y_{i2}, ...)$ with y_{ij} representing the probability that the number of $\mathcal{P}_1$ customers in the system is i and that in $\mathcal{P}_2$ queue is j. Equations (2.4) and (2.5) give

$$
\begin{aligned}
y_0 &= 1 - \rho; \quad \rho = \frac{\lambda}{(\mu_1 + \theta_1)} + \frac{\lambda\, \theta_1}{(\mu_1 + \theta_1)\, \mu_2}, \\
y_{00} &= \frac{1}{\mu_2} (\lambda - r_0 \mu_1)\, y_0, \\
y_{01} &= \frac{1}{\mu_2} \left\{ (\lambda + \mu_2 - r_0 \mu_1)\, y_{00} - (r_0 \theta_1 + r_1 \mu_1)\, y_0 \right\}, \\
y_{0j} &= \tfrac{1}{\mu_2} \Big\{ (\lambda + \mu_2 - r_0\mu_1)\, y_{0,j-1} - \theta_1 \sum_{k=0}^{j-2} r_k y_{0,j-2-k} - \\
&\qquad \mu_1 \sum_{k=1}^{j-1} r_k y_{0,j-1-k} - (r_{j-1}\theta_1 + r_j \mu_1) y_0 \Big\}, \; j > 1.
\end{aligned}
$$

Thus we can compute y_{0j} recursively up to the desired range of values.
Substituting for $\boldsymbol{y}_0$ in equation (2.4) and expanding, $y_{1j}, j = 0, 1, 2, 3...$ are computed as

$$
\begin{aligned}
y_{10} &= (1 - \rho)\, r_0, \\
y_{11} &= (1 - \rho)\, r_1 + y_{00}\, r_0, \\
y_{1j} &= (1 - \rho)\, r_j + \sum_{k=0}^{j-1} y_{0k}\, r_{j-1-k}, \; j = 2, 3, ...
\end{aligned}
$$

Let y_{ij} represent the probability that there are i high priority customers in the system and j low priority customers waiting in queue, $i > 1$. Expression (2.2) on expansion results in

$$
y_{ij} = \sum_{k=0}^{j} y_{i-1,k}\, r_{j-k}, i > 1. \tag{2.9}
$$

After obtaining y_{0j} and y_{1j} for $j = 0, 1, 2, ...,$ $y_{ij}, i > 1$ is recursively computed using (2.9).

2.1.2. *The Marginal Probabilities:* The marginal probabilities of the number of high-priority($\mathcal{P}_1$) customers in the system be denoted by $\mathbf{y}_{i.} = \sum_{j=0}^{\infty} \mathbf{y}_{ij},\ i \geq 0$. So

$$
\mathbf{y}_{i.} = \sum_{j=0}^{\infty} \mathbf{y}_{ij} = \sum_{j=0}^{\infty} \sum_{k=0}^{j} \mathbf{y}_{i-1,k}\, r_{j-k} = \left(\sum_{j=0}^{\infty} \mathbf{y}_{i-1,j} \right) \left(\sum_{i=0}^{\infty} r_i \right) = \mathbf{y}_{(i-1).}\, \rho_1. \tag{2.10}
$$

Remark: As an arrival of a $\mathcal{P}_1$ customer preempts a $\mathcal{P}_2$ customer in service, the system behaves as an M/M/1 queue as far as marginal probabilities of $\mathcal{P}_1$ customers are concerned. Hence

$$\mathbf{y}_{i.} = \rho_1^i(1-\rho_1), i \geq 0; \quad \rho_1 = \frac{\lambda}{\mu_1+\theta_1} \tag{2.11}$$

The marginal distribution of $\mathcal{P}_2$ customers is computed numerically from

$$\mathbf{y}_{.j} = \sum_{i=0}^{\infty} \mathbf{y}_{ij},\ j \geq 0. \tag{2.12}$$

2.2. **Waiting time Analysis.**

Waiting time of High priority customers. As an arriving $\mathcal{P}_1$ customer preempts the $\mathcal{P}_2$ customer if any under service, the waiting time distribution is same as in the case of an M/M/1 queue. Hence expected waiting time of $\mathcal{P}_2$ in the system is

$$E(W_{\mathcal{P}_1}) = \frac{\rho_1}{\lambda\,(1-\rho_1)} = \frac{1}{\mu_1+\theta_1-\lambda}$$

Waiting time of Low priority customers. Expected waiting time of a $\mathcal{P}_2$ customer, provided he is the head of the $\mathcal{P}_2$ line, is the sum of the following: expected busy cycle generated by the primary customers left behind by this customer when he interrupted own service while in $\mathcal{P}_1$, the sum of the expected busy cycles generated at each preemption while chosen for service from $\mathcal{P}_2$ line and expected time taken to complete service without a preemption. We get

$$E(W_{\mathcal{P}_2^1}) = \frac{1}{(\mu_1+\theta_1)}\frac{\rho_1}{(1-\rho_1)^2} + \frac{1}{\mu_2(1-\rho_1)}$$

Expected waiting time of a $\mathcal{P}_2$ customer if he is anywhere in the $\mathcal{P}_2$ line is

$$E(W_{\mathcal{P}_2}) = \frac{1}{\mu_1+\theta_1}\frac{\rho_1}{(1-\rho_1)^2} + \frac{1}{\mu_2}\frac{1}{(1-\rho_1)}E(\mathcal{P}_2);\ E(\mathcal{P}_2) = \sum_{r=1}^{\infty} r y_{\cdot r}$$

3. **Two priority -Non preemptive case**

Model Description: We consider a two-priority queueing model similar to that in the previous section, except that service to the $\mathcal{P}_2$ customers is according to a non-preemptive service discipline. That is the arrival of a $\mathcal{P}_1$ customer does not interrupt the ongoing service of a $\mathcal{P}_2$ customer. We follow the same notations. Let $N_1(t)$ be the number of $\mathcal{P}_1$ customers in the system including the one in service if any, $N_2(t)$ be the number of $\mathcal{P}_2$ waiting to get service and $S(t)$ the status of the server which is 1 or 2 according as the server is busy with $\mathcal{P}_1$ customers or $\mathcal{P}_2$ customers . Thus we get a continuous time Markov chain $\Omega = \{X\,(t)\,, t \geq 0\} = \{(N_1(t), N_2(t), S(t))\,/t \geq 0\}$. Its state space is given as $\{(0,0)\} \cup \{(0,j,2)/j \geq 0\} \cup \{(i,j,k)/i > 0, j \geq 0, k = 1,2\}$.

It is not hard to derive the condition for system stability as $\frac{\lambda}{(\mu_1+\theta_1)} + \frac{\lambda\,\theta_1}{(\mu_1+\theta_1)\mu_2} < 1,$

The infinitesimal generator of this continuous time Markov chain consists of block entries of infinite dimension and is obtained as

$$Q^* = \begin{pmatrix} A_{00} & A_{01} & & & \\ A_{10} & A_1 & A_0 & & \\ & A_2 & A_1 & A_0 & \\ & & \ddots & \ddots & \ddots \end{pmatrix}$$

where,

$$A_{00} = \begin{pmatrix} -\lambda & & \\ \mu_2 & -(\lambda+\mu_2) & \\ & \mu_2 & -(\lambda+\mu_2) \\ & \ddots & \ddots \end{pmatrix}, A_{01} = \begin{pmatrix} [\lambda] & & & & \\ & [\lambda \quad 0] & & & \\ & & [\lambda \quad 0] & & \\ & & & [\lambda \quad 0] & \\ & & & & \ddots \end{pmatrix},$$

$$A_{10} = \begin{pmatrix} \begin{bmatrix} \mu_1 \\ 0 \end{bmatrix} & \begin{bmatrix} \theta_1 & 0 \\ 0 & 0 \end{bmatrix} & & \\ & \begin{bmatrix} \mu_1 & \theta_1 \\ 0 & 0 \end{bmatrix} & & \\ & & \begin{bmatrix} \mu_1 & \theta_1 \\ 0 & 0 \end{bmatrix} & \\ & & & \ddots \end{pmatrix},$$

$$A_2 = \begin{pmatrix} M_2 & M_3 & & \\ & M_2 & M_3 & \\ & & \ddots & \ddots \end{pmatrix}, \ A_1 = \begin{pmatrix} M_1 & & \\ & M_1 & \\ & & \ddots \end{pmatrix}, \ A_0 = \lambda I_\infty;$$

$$M_1 = \begin{bmatrix} -(\lambda+\mu_1+\theta_1) & 0 \\ \mu_2 & -(\lambda+\mu_2) \end{bmatrix}, M_2 = \begin{bmatrix} \mu_1 & 0 \\ 0 & 0 \end{bmatrix}, M_3 = \begin{bmatrix} \theta_1 & 0 \\ 0 & 0 \end{bmatrix}.$$

The infinitesimal generator Q^* constitutes a quasi birth and death(QBD) process with infinite number of sub-levels. As Q^* is irreducible and recurrent, following a similar argument to theorem 3 of Miller [7] we have,

Theorem 3.1. *Let $\boldsymbol{x} = [\boldsymbol{x}_0, \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots]$ denote the invariant probability vector for the QBD process Q^* with infinite number of sub levels(phases), where $\boldsymbol{x}_i$ is the probability vector corresponding to level i of infinite dimension. Then the solution for $\boldsymbol{x}$ possesses a matrix geometric structure*

$$\boldsymbol{x}_i = \boldsymbol{x}_{i-1} R, \ i > 1. \tag{3.1}$$

where the rate matrix R is the minimal non negative solution to

$$R^2 A_2 + R A_1 + A_0 = 0. \tag{3.2}$$

Theorem 3.2. *The R matrix, which is the minimal non negative solution to equation (3.2) possesses a Toeplitz structure $(R_0, R_1, R_2, ...)$. That is R has the form*

$$R = \begin{pmatrix} R_0 & R_1 & R_2 & R_3 & \dots \\ 0 & R_0 & R_1 & R_2 & \dots \\ 0 & 0 & R_0 & R_1 & \dots \\ 0 & 0 & 0 & R_0 & \dots \\ \dots & \dots & \dots & \dots & \dots \end{pmatrix}$$

where each of the matrices R_k is of order 2 represented as $R_k = \begin{bmatrix} a_k & 0 \\ b_k & c_k \end{bmatrix}$.

Proof. The interpretation of R in Neuts [8] and the structure of the matrices in the generator matrix Q proves the theorem. □

Theorem 3.3. *The elements $R_k(k > 0)$ in theorem 3.2 are computed as,*

$$a_k = \frac{(\sum_{i=o}^{k-1} a_i a_{k-1-i})\theta + (\sum_{i=1}^{k-1} a_i a_{k-i})\mu_1}{(\lambda + \mu_1 + \theta) - 2a_0\mu_1},$$

$$b_k = \frac{(\sum_{i=1}^{k-1} a_i b_{k-1-i} + b_{k-1}(a_0 + c_0))\theta + (\sum_{i=1}^{k} a_i b_{k-i})\mu_1}{(\lambda + \mu_1 + \theta) - (a_0 + c_0)\mu_1},$$

$$c_k = 0, \qquad k = 1, 2, 3, \dots$$

and entries of R_0 are

$$a_0 = \frac{(\lambda + \mu_1 + \theta) - \sqrt{(\lambda + \mu_1 + \theta)^2 - 4\mu_1\lambda}}{2\mu_1},$$

$$b_0 = \frac{\mu_2 c_0}{(\lambda + \mu_1 + \theta) - (a_0 + c_0)\mu_1},$$

$$c_0 = \frac{\lambda}{\lambda + \mu_2}.$$

Proof. Upon expansion of (3.2), we obtain the following relations:

$$R_0^2 M_2 + R_0 M_1 + \lambda I = 0, \tag{3.3}$$

$$R_0^2 M_3 + \left(\sum_{k=0}^{l} R_k R_{l-k} \right) M_2 + R_l M_1 = 0, \text{ for } l \geq 1. \tag{3.4}$$

The result is established when these equations are expanded with respect to the phases. □

3.1. **The joint and marginal probabilities.** In this section, the recursive formulas for the joint distribution of i number of $\mathcal{P}_1$ customers in the system and j number of $\mathcal{P}_2$ customers in the queue and marginal distributions of each are derived. First we establish the following.

Theorem 3.4. *The matrix geometric structure*

$$\boldsymbol{x}_i = \boldsymbol{x}_{i-1}R, \quad i > 1$$

given in theorem 3.1 extended to level 0 is

$$\boldsymbol{x}_i = \boldsymbol{x}_0 \left(\frac{1}{\lambda}A_{01}\right) R^i, \quad i \geq 1.$$

Proof. From $\boldsymbol{x}Q^* = 0$, the two boundary equations involving $\boldsymbol{x}_0$ are

$$\boldsymbol{x}_0 A_{00} + \boldsymbol{x}_1 A_{10} = 0, \tag{3.5}$$

$$\boldsymbol{x}_0 A_{01} + \boldsymbol{x}_1 [A_1 + RA_2] = 0. \tag{3.6}$$

From (3.2) it follows that

$$R[RA_2 + A_1] = -A_0. \tag{3.7}$$

Since $A_0 = \lambda I_\infty$, R is invertible.
From (3.6) and (3.7) we get

$$\boldsymbol{x}_1 = \boldsymbol{x}_0 \left(\frac{1}{\lambda}A_{01}R\right). \tag{3.8}$$

Combining relations (3.1) and (3.8) we obtain

$$\boldsymbol{x}_i = \boldsymbol{x}_0 \left(\frac{1}{\lambda}A_{01}\right) R^i, \quad i \geq 1. \tag{3.9}$$

□

3.1.1. *The Joint Probability Distribution.* Let $\boldsymbol{x}_{ij}$ be the probability that there are i high priority customers in the system and j low priority customers waiting in queue. Further let the marginal distribution of the number of high-priority customers in system be denoted by

$$\boldsymbol{x}_{i.} = \sum_{j=0}^{\infty} \boldsymbol{x}_{ij} \ , \quad i \geq 0. \tag{3.10}$$

To get to know the type of customer in service we partition $\boldsymbol{x}_{ij}$ as

$$\boldsymbol{x}_{ij} = (x_{ij}(1), x_{ij}(2)). \tag{3.11}$$

We proceed to the determination of the joint probability vectors $\boldsymbol{x}_{ij}$. Considering the interrupted customers and the type of customer under service, equation (3.1) gives

$$\boldsymbol{x}_{ij} = \boldsymbol{x}_{i-1,j}R \ , \quad i > 1, j \geq 0. \tag{3.12}$$

where

$$\boldsymbol{x}_{ij} = (\boldsymbol{x}_{ij}(1), \boldsymbol{x}_{ij}(2)). \tag{3.13}$$

Expanding (3.12) w.r.t. j

$$\begin{pmatrix} \boldsymbol{x}_{i0}, & \boldsymbol{x}_{i1}, & \cdots \end{pmatrix} = \begin{pmatrix} \boldsymbol{x}_{i-1,0}, & \boldsymbol{x}_{i-1,1}, & \cdots \end{pmatrix} \times \begin{pmatrix} R_0 & R_1 & R_2 & \cdots \\ 0 & R_0 & R_1 & \cdots \\ 0 & 0 & R_0 & \cdots \\ 0 & 0 & 0 & \ddots \end{pmatrix}. \tag{3.14}$$

In general,

$$\boldsymbol{x}_{ij} = \sum_{k=0}^{j} \boldsymbol{x}_{i-1,k}\ R_{j-k}\ , \quad i > 1, j \geq 0. \tag{3.15}$$

Expanding these equations once more to reveal the dependence on the type of service, we obtain

$$\boldsymbol{x}_{ij}(1) = \sum_{k=0}^{j} [a_{j-k}\ \boldsymbol{x}_{i-1,k}(1) + b_{j-k}\ \boldsymbol{x}_{i-1,k}(2)] \tag{3.16}$$

$$\boldsymbol{x}_{ij}(2) = c_0\ \boldsymbol{x}_{i-1,j}(2); \qquad i > 1, j \geq 0. \tag{3.17}$$

Equation (3.8) on expansion gives

$$\boldsymbol{x}_{1j}(1) = a_j(1-\rho) + \sum_{k=0}^{j} b_{j-k}\ \boldsymbol{x}_{0k}(2) \tag{3.18}$$

$$\boldsymbol{x}_{1j}(2) = c_0\ \boldsymbol{x}_{0j}(2); \qquad i = 1, j \geq 0. \tag{3.19}$$

Hence the joint probabilities depend on $\boldsymbol{x}_{0k}(2)$ for $k = 0, 1, 2, \ldots j$. We compute $\boldsymbol{x}_{0k}(2)$ in the desired range in the next section.

3.1.2. *Marginal Distribution of High Priority customers.* Adding equation(3.15) over j, the low priority queue length, the marginal distribution $\boldsymbol{x}_{i.}$ for the number of high priority customers in the system is

$$\begin{aligned} \boldsymbol{x}_{i.} &= \sum_{j=0}^{\infty} \boldsymbol{x}_{ij} = \sum_{j=0}^{\infty}\sum_{k=0}^{j} \boldsymbol{x}_{i-1,k}\ R_{j-k} = \sum_{k=0}^{\infty} \boldsymbol{x}_{i-1,k} \left(\sum_{j=0}^{\infty} R_j \right) \\ &= \boldsymbol{x}_{(i-1).} \mathcal{R}_+ \qquad (3.20) \\ &= \boldsymbol{x}_{1.} \mathcal{R}_+^{i-1}, \quad i \geq 2; \qquad (3.21) \end{aligned}$$

where

$$\mathcal{R}_+ = \sum_{j=0}^{\infty} R_j = \begin{bmatrix} \sum_{r=0}^{\infty} a_r & 0 \\ \sum_{r=0}^{\infty} b_r & c_0 \end{bmatrix}.$$

Now, expanding (3.20) based on the type of service, we have

$$\begin{aligned}\left(\begin{array}{cc} \boldsymbol{x}_i(1), & \boldsymbol{x}_i(2) \end{array}\right) &= \left(\begin{array}{cc} \boldsymbol{x}_{i-1}(1), & \boldsymbol{x}_{i-1}(2) \end{array}\right) \begin{bmatrix} \sum a_r & 0 \\ \sum b_r & c_0 \end{bmatrix}, i \geq 1 \\ &= \left(\begin{array}{cc} \boldsymbol{x}_{i-1}(1)\,(\sum a_r) + \boldsymbol{x}_{i-1}(2)\,(\sum b_r), & \boldsymbol{x}_{i-1}(2)\ c_0 \end{array}\right)\end{aligned}$$

So we obtain

$$\begin{aligned}\boldsymbol{x}_{i.}(1) &= \boldsymbol{x}_{i-1,.}(1)\,(\sum a_r) + \boldsymbol{x}_{i-1,.}(2)\,(\sum b_r) \\ \boldsymbol{x}_{i.}(2) &= \boldsymbol{x}_{i-1,.}(2)\ c_0\end{aligned}$$

Adding equations (3.18) and (3.19) over j

$$\boldsymbol{x}_{1.}(1) = (1-\rho)\left(\sum a_r\right) + \boldsymbol{x}_{0.}(2)\left(\sum b_r\right) \tag{3.22}$$

$$\boldsymbol{x}_{1.}(2) = c_0\ \boldsymbol{x}_{0.}(2) \tag{3.23}$$

which in turn gives

$$\boldsymbol{x}_{1.} = \left(\begin{array}{cc} (1-\rho), & \boldsymbol{x}_{0.}(2) \end{array}\right)\mathcal{R}_+ \tag{3.24}$$

Combining equations (3.21) and (3.24) we get

$$\boldsymbol{x}_{i.} = \left(\begin{array}{cc} (1-\rho), & \boldsymbol{x}_{0.}(2) \end{array}\right)\mathcal{R}_+^i \quad ; \quad i \geq 1. \tag{3.25}$$

Write $\boldsymbol{x}_{0.} = \left(\begin{array}{cc} (1-\rho), & \boldsymbol{x}_{0.}(2) \end{array}\right)$, then

$$\boldsymbol{x}_{i.} = \boldsymbol{x}_{0.}\mathcal{R}_+^i \quad ; \quad i \geq 1. \tag{3.26}$$

On expansion of equation (3.26) we get the high priority marginals as

$$\boldsymbol{x}_{i.}(1) = (1-\rho)\left(\sum a_r\right)^i + \boldsymbol{x}_{0.}(2)\sum_{k=0}^{i-1}\left(\sum a_r\right)^k\left(\sum b_r\right)c_0^{i-1-k} \tag{3.27}$$

$$\boldsymbol{x}_{i.}(2) = \boldsymbol{x}_{0.}(2)c_0^i \tag{3.28}$$

From equations (3.27) and (3.28) it is clear that the marginal probabilities depend on the probability that no $\mathcal{P}_1$ customer and a P_2 customer in service, which is given by

$$\boldsymbol{x}_{0.}(2) = \sum_{j=0}^{\infty}\boldsymbol{x}_{0j}(2). \tag{3.29}$$

To compute $\boldsymbol{x}_{0.}(2)$*:* Substituting equation (3.8) in (3.5)

$$\boldsymbol{x}_0\left[A_{00} + \frac{1}{\lambda}(A_{01}RA_{10})\right] = 0 \tag{3.30}$$

$\boldsymbol{x}_0$ being given by $\boldsymbol{x}_0 = \left(\begin{array}{cccccc} (1-\rho), & \boldsymbol{x}_{00}(2), & \boldsymbol{x}_{01}(2), & \boldsymbol{x}_{02}(2)\ , & \ldots & \ldots \end{array}\right).$

Expanding(3.30), the following relations are obtained:

$$(1-\rho)\left[a_0\mu_1-\lambda\right]+\boldsymbol{x}_{00}(2)\left[b_0\mu_1+\mu_2\right]=0 \tag{3.31}$$

$$\begin{aligned}(1-\rho)\left[a_0\theta_1+a_1\mu_1\right]+\boldsymbol{x}_{00}(2)\left[b_0\theta_1+b_1\mu_1-(\lambda+\mu_2)\right]&\\ +\boldsymbol{x}_{01}(2)\left[b_0\mu_1+\mu_2\right]&=0 \end{aligned}\tag{3.32}$$

For $j\geq 2$;

$$\begin{aligned}(1-\rho)\left[a_{j-1}\theta_1+a_j\mu_1\right]+\sum_{k=0}^{j-2}\boldsymbol{x}_{0k}(2)\left[b_{j-k-1}\theta_1+b_{j-k}\mu_1\right]&\\ +\boldsymbol{x}_{0(j-1)}(2)\left[b_0\theta_1+b_1\mu_1-(\lambda+\mu_2)\right]+\boldsymbol{x}_{0j}(2)\left[b_0\mu_1+\mu_2\right]&=0.\end{aligned}\tag{3.33}$$

On solving these, the following are obtained.

$$\boldsymbol{x}_{00}(2)=\frac{(\lambda-a_0\mu_1)(1-\rho)}{b_0\mu_1+\mu_2} \tag{3.34}$$

(3.35)
$$\boldsymbol{x}_{01}(2)=\frac{1}{b_0\mu_1+\mu_2}\left\{\left[(\lambda+\mu_2)-(b_0\theta_1+b_1\mu_1)\right]\boldsymbol{x}_{00}(2)-(1-\rho)\left[a_0\theta_1+a_1\mu_1\right]\right\}$$

$$\begin{aligned}\boldsymbol{x}_{0j}(2)=\frac{1}{b_0\mu_1+\mu_2}\Bigg\{&\left[(\lambda+\mu_2)-(b_0\theta_1+b_1\mu_1)\right]\boldsymbol{x}_{0(j-1)}(2)-(1-\rho)\left[a_{j-1}\theta_1+a_j\mu_1\right]\\ &-\sum_{k=0}^{j-2}\boldsymbol{x}_{0k}(2)\left[b_{j-k-1}\theta_1+b_{j-k}\mu_1\right]\Bigg\},\ j\geq 2.\end{aligned}\tag{3.36}$$

Hence $\boldsymbol{x}_{0.}(2)$ in equation(3.29) is computed. Also the joint probabilities given by relations (3.16) to (3.19) are evaluated.

3.1.3. *Marginal Distribution of Low Priority customers.* Define $\boldsymbol{x}_{.j}(1)=\sum_{i=1}^{\infty}\boldsymbol{x}_{ij}(1)$ and $\boldsymbol{x}_{.j}(2)=\sum_{i=0}^{\infty}\boldsymbol{x}_{ij}(2)$ for $j\geq 0$.

Summing equations (3.16) from $i=2$ to ∞ and adding this to (3.18) we obtain

(3.37)
$$\begin{aligned}\boldsymbol{x}_{.j}(1) &= a_j(1-\rho)+\textstyle\sum_{i=2}^{\infty}\sum_{k=0}^{j}a_{j-k}\,\boldsymbol{x}_{(i-1)k}(1)+\sum_{i=1}^{\infty}\sum_{k=0}^{j}b_{j-k}\,\boldsymbol{x}_{(i-1)k}(2)\\ &= a_j(1-\rho)+\textstyle\sum_{k=0}^{j}\left[a_{j-k}\,\boldsymbol{x}_{.k}(1)+b_{j-k}\,\boldsymbol{x}_{.k}(2)\right]\end{aligned}$$

Similarly adding equations (3.17) from $i=2$ to ∞ and adding this to (3.19),

$$\begin{aligned}\boldsymbol{x}_{.j}(2)&=\textstyle\sum_{i=0}^{\infty}\boldsymbol{x}_{ij}(2)\\ &=\boldsymbol{x}_{0j}(2)+\textstyle\sum_{i=1}^{\infty}\boldsymbol{x}_{ij}(2)\\ &=\boldsymbol{x}_{0j}(2)+c_0\textstyle\sum_{i=1}^{\infty}\boldsymbol{x}_{(i-1)j}(2)\\ &=\boldsymbol{x}_{0j}(2)+c_0\,\boldsymbol{x}_{.j}(2).\end{aligned}$$

$$\implies \qquad \boldsymbol{x}_{.j}(2)=\frac{1}{1-c_0}\,\boldsymbol{x}_{0j}(2). \tag{3.38}$$

Hence the marginal probabilities of low priority customers while a $\mathcal{P}_2$ customer is under service, is determined once we determine $x_{0j}(2)$ for the desired range of values of j, which is done through equations (3.34) and (3.36). The marginal probabilities of low priority customers, while a $\mathcal{P}_1$ customer is under service, is determined as follows. Substituting for $\boldsymbol{x}_{0j}(2)$ and putting $k = 0, 1, 2, ..., j$ in (3.37) we get

$$\boldsymbol{x}_{.0}(1) = \frac{a_0\,(1-\rho) + b_0\,\boldsymbol{x}_{.0}(2)}{1-a_0} \tag{3.39}$$

$$\boldsymbol{x}_{.j}(1) = \frac{a_j\,(1-\rho) + \sum_{k=0}^{j-1} a_{j-k}\,\boldsymbol{x}_{.k}(1) + \sum_{k=0}^{j} b_{j-k}\,\boldsymbol{x}_{.k}(2)}{1-a_0}, \quad j \geq 1 \tag{3.40}$$

3.2. Waiting time distribution.

3.2.1. *High priority waiting time distribution.* First we compute the expected waiting time of a $\mathcal{P}_1$ customer who joins as the n^{th} customer $n(>0)$, in the queue at the time when he joins. We construct a Markov chain $\{N(t), t \geq 0\}$, where $N(t)$ is the rank of the customer at time t. The rank of a customer is r if he is the r^{th} customer in the queue at time t. His rank decreases by 1 as the customers ahead of him leave the system after completing/ self interrupting service. Two cases are to be considered according as whether a $\mathcal{P}_1$ or a $\mathcal{P}_2$ customer is under service when the tagged customer joins.

State space of the Markov chain when a $\mathcal{P}_1$ customer is in service is $\{n : 1 \leq n < r\} \cup \{(r,1)\} \cup \{0\}$ and that when a $\mathcal{P}_2$ customer is in service is $\{n : 1 \leq n < r\} \cup \{(r,2)\} \cup \{0\}$, where $\{0\}$ is the absorbing state indicating that the tagged customer is selected for service. The corresponding infinitesimal generator matrices of dimension $r+1$ are denoted by $\mathcal{W}_1$ and $\mathcal{W}_2$ respectively, and are

$$\mathcal{W}_1 = \begin{bmatrix} T_r & T_r^0 \\ 0 & 0 \end{bmatrix}, \mathcal{W}_2 = \begin{bmatrix} S_r & S_r^0 \\ 0 & 0 \end{bmatrix} \text{ where,}$$

$$T_r = \begin{cases} -\mu_1\,,\ i = j = 1, 2, ..., r. \\ \mu_1\,,\ j = i+1, i = 1, 2, ..., r-1 \\ 0\,,\ elsewhere \end{cases}, S_r = \begin{cases} -\mu_2\,, i = j = 1 \\ \mu_2\,, i = 1, j = 2 \\ -\mu_1\,,\ i = j = 1, 2, ..., r. \\ \mu_1\,,\ j = i+1, i = 1, 2, ..., r-1 \\ 0\,,\ elsewhere \end{cases}$$

and $T_r^0 = S_r^0 = [0\ ...\ 0\ \mu_1]^T$.

If $\mathbf{e}$ is a column vector of ones of appropriate order, then the expected waiting time of the r^{th} tagged customer is $-(T_r^{-1} + S_r^{-1})\mathbf{e}$.

Hence the expected waiting time of a $\mathcal{P}_1$ customer in the queue, with $\boldsymbol{\alpha} = [1\ 0\ ...\ 0]$ a row vector of dimension r is,

$$W_{\mathcal{P}_1} = \sum_{r=1}^{\infty} \left[\left(-\alpha T_r^{-1} e\right)\ x_{(r+1).}(1) + \left(-\alpha S_r^{-1} e\right) x_{r.}(2) \right]$$

3.2.2. *Low priority waiting time distribution.* We compute the bounds on the distribution of waiting time of an arriving(tagged) customer in the system. When the tagged customer arrives either a $\mathcal{P}_1$ or a $\mathcal{P}_2$ will be under service or the server is free. Suppose the tagged customer joins as r^{th} in the system. The probability of observing these events are $\boldsymbol{x}_{..}(2)$, $\boldsymbol{x}_{..}(1)$ and $1 - \rho$.

Service time distribution $\mathbf{S}_1$ of the customer under service if it is a $\mathcal{P}_2 \sim exp(\mu_2)$ and service time distribution $\mathbf{S}_2^{'}$ of the customer under service if it is a $\mathcal{P}_1 \sim exp(\mu_1)$. But $\mathbf{S}_2^{'}$ is clubbed with the distribution of service time of $\mathcal{P}_1$ customers in the queue since his service is followed by customers if any, in the $\mathcal{P}_1$ line.

The waiting time distribution of these r $\mathcal{P}_1$ customers is the $r-$ fold convolution of $exp(\mu_1)$, that is $\mathbf{E}(r, \mu_1)$. Hence the distribution of service(wait) time until the tagged customer interrupts own service

$$\mathbf{S}_2 = \sum_{r=2}^{\infty} \mathbf{E}(r, \mu_1)\ \boldsymbol{x}_{(r-1).}(1) + exp(\mu_1).(1 - \rho)$$

So, the distribution of waiting time of a customer in the system until interruption is

$$\mathbf{F}_0 = \boldsymbol{x}_{..}(2)\mathbf{S}_1 * \mathbf{S}_2$$

Now assume that the tagged customer interrupts. Probability to interrupt is θ_1. We may assume without loss of generality, that the tagged customer leave behind 'i' $\mathcal{P}_1$ customers at his service interruption and join as j^{th} in the $\mathcal{P}_2$ line. Each of these i $\mathcal{P}_1$ customers generate a busy cycle exponentially distributed with parameter $(\mu_1 + \theta_1 - \lambda)$. So their service time is $i-$fold convolution of $exp(\mu_1 + \theta_1 - \lambda)$ with itself. The probability to see i customers behind the tagged customer in $\mathcal{P}_1$ line is $\boldsymbol{x}_{i.}(1)$. Thus the distribution of service time of these i customers is

$$\mathbf{F}_1 = \sum_{i=0}^{\infty} \mathbf{E}(i, \mu_1 + \theta_1 - \lambda)\ \boldsymbol{x}_{i.}(1)$$

where $\mathbf{E}(i, \alpha)$ stands for Erlang distribution of order i and parameter α. When $i = 0$, that is when no $\mathcal{P}_1$ behind him at the time of interruption $\boldsymbol{x}_{i.}(1) = 0$.

The lower bound.

The waiting time of the tagged customer is minimum if no $\mathcal{P}_1$ customers arrive during the service of all the $(j - 1)$ customers ahead of the tagged customer in $\mathcal{P}_2$ line. Now, suppose that once the server is at the $\mathcal{P}_2$ line it returns to $\mathcal{P}_1$ line only after serving the tagged $\mathcal{P}_2$(that is, all the $j-1$ $\mathcal{P}_2$ and tagged $\mathcal{P}_2$ complete service in a row). This is possible if no $\mathcal{P}_1$ customer arrives once the service in $\mathcal{P}_2$ line started.

The probability of finding $(j-1)$ $\mathcal{P}_2$ ahead of tagged $\mathcal{P}_2$,

$$q_j^{'} = \boldsymbol{x}_{0(j-2)}(2) + \boldsymbol{x}_{.(j-1)}(1)$$

The probability that no $\mathcal{P}_1$ arrived during the service time of a $\mathcal{P}_2$ customer, which is exponentially distributed with μ_2 is

$$p_0 = \int_0^\infty \frac{e^{-\lambda t}(\lambda t)^0}{0!} \mu_2 e^{-\mu_2 t} dt = \int_0^\infty e^{-\lambda t} \mu_2 e^{-\mu_2 t} dt$$

Therefore the probability that no $\mathcal{P}_1$ arrived during their($(j-1)$ $\mathcal{P}_2$) service time is

$$q_{j-1} = p_0^{j-1}$$

Note: *When $j = 1$, $q_0 = 1$ indicating that if tagged $\mathcal{P}_2$ is the head of the $\mathcal{P}_2$ line, then he is taken for service (prob.=1) immediately(no $\mathcal{P}_1$ arrival when no $\mathcal{P}_2$ ahead of tagged $\mathcal{P}_2$) when the server is at $\mathcal{P}_2$ line.*

Therefore the corresponding service time distribution is $j-$fold convolution of $exp(\mu_2)$ with itself multiplied by the probabilities q_j^* and q_{j-1}. Therefore the service time of j $\mathcal{P}_2$ is

$$\mathbf{F}_2 = \sum_{j=1}^{\infty} \mathbf{E}(j, \mu_2)\, q_j^{'}\, q_{j-1}.$$

So we get the distribution of the lower bound of waiting time in the system as the convolution

$$\mathbf{F}_{\min wait} = \mathbf{F}_0 * \theta_1 \mathbf{F}_1 * \mathbf{F}_2.$$

The upper bound.

The waiting time of the tagged customer is maximum if $\mathcal{P}_1$ customers arrive during the service of each of $(j-1)$ customers ahead of the tagged customer in $\mathcal{P}_2$ line. Hence immediately after the service of each $\mathcal{P}_2$ the server goes to $\mathcal{P}_1$ line and returns to $\mathcal{P}_2$ line till the tagged $\mathcal{P}_2$ get service. Suppose k $\mathcal{P}_1$ customers lined up during the service of a $\mathcal{P}_2$. The probability that k $\mathcal{P}_1$ arrived during the service time of one $\mathcal{P}_2$ customer, which is exponentially distributed with μ_2 is

$$p_k = \int_0^\infty \frac{e^{-\lambda t}(\lambda t)^k}{k!} \mu_2 e^{-\mu_2 t} dt$$

Distribution of waiting time(service time) due to these k $\mathcal{P}_2$ is $k-$fold convolution of $exp(\mu_1 + \theta_1 - \lambda)$ with itself. Therefore distribution of waiting time after the service of each $\mathcal{P}_2$ from among the $j-1$ $\mathcal{P}_2$ is

$$\sum_{k=1}^{\infty} E(k, \mu_1 + \theta_1 - \lambda)\, p_k$$

Waiting time distribution generated by the service of all $j-1$ $\mathcal{P}_2$ ahead of the tagged customer is

$$\mathbf{F}_3 = \sum_{j=1}^{\infty} \left[\exp(\mu_2) * \sum_{k=1}^{\infty} \mathbf{E}(k, \mu_1 + \theta_1 - \lambda)\ p_k \right]^{*(j-1)} q_j^{'},$$

$*$ stands for the convolution with $*r$ denoting the $r-$ fold convolution. Hence the distribution of Maximum waiting time of a customer in the system if he interrupts own service is

$$\mathbf{F}_{\max wait} = \mathbf{F}_0 * \theta_1\ \mathbf{F}_1 * \mathbf{F}_3 *\ \exp(\mu_2).$$

3.3. **Additional Performance Measures and their numerical illustrations.**

1. The probability that all the $\mathcal{P}_1$ customers served in a given cycle complete service without any interruption
$$P_{AC} = \frac{\mu_1(\mu_1 + \theta_1 - \lambda)}{(\mu_1 + \theta_1)^2 - \lambda\mu_1}.$$
This is equivalent to seeking the probability that there is no inflow to $\mathcal{P}_2$ from $\mathcal{P}_1$ during that cycle.
2. The probability that all the $\mathcal{P}_1$ customers served in a given cycle interrupt before completing service and hence go to $\mathcal{P}_2$
$$P_{AI} = \frac{\theta_1(\mu_1 + \theta_1 - \lambda)}{(\mu_1 + \theta_1)^2 - \lambda\theta_1}$$
This is the probability for the other extreme case of 1.

We demonstrate below the impact of fixed values of λ, μ_1, andμ_2 on P_{AC} and P_{AI} with variations of θ_1. In tables 1 and 2, P_{AC} and P_{AI} have identical values corresponding to $\theta_1 = \mu_1 = 6$.

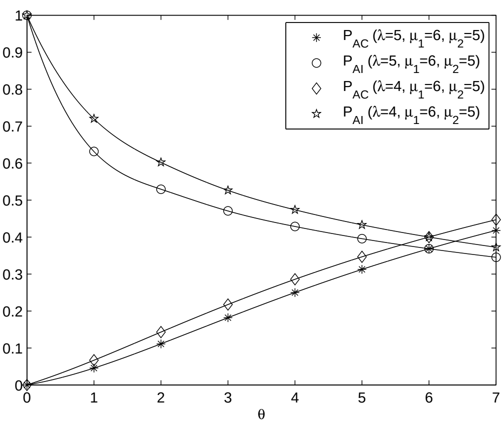

FIGURE 1

The figure clearly shows that as the value of θ_1 increases P_{AC} decreases and P_{AI} increases.

θ_1	0	1	2	3	4	5	6	7
P_{AC}	1	.6316	.5294	.4706	.4286	.3954	**.3684**	.3453
P_{AI}	0	.0455	.1111	.1818	.2500	.3125	**.3684**	.4179

TABLE 1. $\lambda = 5, \mu_1 = 6, \mu_2 = 5$

θ_1	0	1	2	3	4	5	6	7
P_{AC}	1	.7200	.6000	.5263	.4737	.4330	**.4000**	.3724
P_{AI}	0	.0667	.1429	.2174	.2857	.3465	**.4000**	.4468

TABLE 2. $\lambda = 4, \mu_1 = 6, \mu_2 = 5$

4. Case of three priorities, non-preemptive:

Model Description: In this section we consider a single server infinite capacity queuing system in which customers from outside arrive according to a Poisson process with rate λ and form a queue($\mathcal{P}_1$) if server is busy. Service times are exponentially distributed with parameter μ_1. Customers in primary queue interrupt service according to a Poisson process of rate θ_1, in which case he has to go to the lower priority queue($\mathcal{P}_2$). Else, he completes service and leaves the system forever. $\mathcal{P}_2$ customers are taken for service according to head of the line priority whenever the queue of external customers is found to be empty at a service completion epoch. The service of such customers is according to a non-preemptive service discipline and the service times are independent and identically distributed exponential random variables with parameter μ_2. A customer from $\mathcal{P}_2$ queue may interrupt his service according to a Poisson process of rate θ_2, up on which he has to go to a third waiting line $\mathcal{P}_3$ (of infinite capacity) and wait for his turn for service. The service time of customers in the third queue are independent and identically distributed exponential random variables with parameter μ_3. Their service is also according to non-preemptive service discipline and customers leave the system after completing service without further interruption. When the server is in $\mathcal{P}_3$ line, $\mathcal{P}_2$ line will be empty whereas in $\mathcal{P}_1$ there may be none, one or more customers.

Let $N_1(t)$ be the number of $\mathcal{P}_1$ customers in the system, $N_j(t)$ that of $\mathcal{P}_j$ customers in the queue for $j = 2, 3$; $S(t)$ the status of the server which is 1, 2 or 3 according as the server is busy with a $\mathcal{P}_1$, $\mathcal{P}_2$ or $\mathcal{P}_3$ customer respectively. Then $\Omega = \{(N_1(t), N_2(t), N_3(t), S(t)) / t \geq 0\}$ is a CTMC with state space $\{0\} \bigcup$
$\{(0, n_2, n_3, k)/n_2 \geq 0, n_3 \geq 0, k = 2, 3\} \bigcup$
$\{(n_1, n_2, n_3, k)/n_1 > 0, n_2 \geq 0, n_3 \geq 0, k = 1, 2, 3\}$.

The condition for stability of the system is given by

$$\frac{\lambda}{(\mu_1 + \theta_1)} + \frac{\lambda\theta_1}{(\mu_1 + \theta_1)(\mu_2 + \theta_2)} + \frac{\lambda\theta_1\theta_2}{(\mu_1 + \theta_1)(\mu_2 + \theta_2)\mu_3} < 1.$$

The infinitesimal generator is obtained as

$$Q = \begin{pmatrix} A_{00}^{(3)} & A_{01}^{(3)} & & & \\ A_{10}^{(3)} & A_{1}^{(3)} & A_{0}^{(3)} & & \\ & A_{2}^{(3)} & A_{1}^{(3)} & A_{0}^{(3)} & \\ & & \ddots & \ddots & \ddots \end{pmatrix}$$

where,

$$A_0^{(3)} = \lambda I_\infty$$

$$A_1^{(3)} = I_\infty \otimes H_3, H_3 = \begin{pmatrix} L_3 & U_3^{(2)} & & & & \\ & L_3 & U_3^{(2)} & & & \\ & & L_3 & U_3^{(2)} & & \\ & & & \ddots & \ddots & \\ & & & & \ddots & \ddots \end{pmatrix}$$

$$\dim(L_3) = 3, \dim(U_3^{(2)}) = 3$$

$$(L_3)_{ij} = \begin{cases} -(\lambda + \mu_i + \theta_i) & ; i = j = 1, 2. \\ -(\lambda + \mu_3) & ; i = j = 3. \\ \mu_i & ; j = 1, i = 2, 3. \\ 0 & ; \text{otherwise} \end{cases}, (U_3^{(2)})_{ij} = \begin{cases} \theta_2 & ; i = 2, j = 1 \\ 0 & ; \text{otherwise} \end{cases}$$

$$A_2^{(3)} = \begin{pmatrix} I_\infty \otimes M_3 & I_\infty \otimes N_3 & & & \\ & I_\infty \otimes M_3 & I_\infty \otimes N_3 & & \\ & & \ddots & \ddots & \\ & & & \ddots & \ddots \\ & & & & \end{pmatrix}$$

$$\dim(M_3) = 3, \dim(N_3) = 3$$

$$(M_3)_{ij} = \begin{cases} \mu_1 & ; i = j = 1 \\ 0 & ; \text{otherwise} \end{cases}, (N_3)_{ij} = \begin{cases} \theta_1 & ; i = j = 1 \\ 0 & ; \text{otherwise} \end{cases}$$

$$A_{01}^{(3)} = \begin{pmatrix} K_3^{(0)} & & & \\ I_\infty \otimes K_3 & & & \\ & I_\infty \otimes K_3 & & \\ & & I_\infty \otimes K_3 & \\ & & & \ddots \end{pmatrix},$$

$$K_3^{(0)} = \begin{pmatrix} \lambda & 0 & 0 & ... \end{pmatrix}, dim(K_3) = 2\times 3, (K_3)_{ij} = \begin{cases} \lambda\,; & j=i+1, i=1,2. \\ 0\,; & \text{elsewhere.} \end{cases}$$

$$A_{10}^{(3)} = \begin{pmatrix} C_3^* & C_3^{(0)} + I_\infty \otimes C_3^{(1)} & & & \\ & I_\infty \otimes C_3^{(2)} & I_\infty \otimes C_3^{(1)} & & \\ & & I_\infty \otimes C_3^{(2)} & I_\infty \otimes C_3^{(1)} & \\ & & & \ddots & \ddots \\ & & & & \ddots \\ & & & & \end{pmatrix}, C_3^* = \begin{pmatrix} \mu_1 \\ 0 \\ 0 \\ \vdots \\ \end{pmatrix}$$

$$C_3^{(0)} = \begin{pmatrix} 0 & & & & \\ B_1 & & & & \\ & B_1 & & & \\ & & B_1 & & \\ & & & \ddots & \end{pmatrix},\ \dim(B_1) = 3\times 2, (B_1)_{ij} = \begin{cases} \mu_1\,; & i=1, j=2 \\ 0\,; & \text{elsewhere} \end{cases}$$

$$\dim(C_3^{(1)}) = \dim(C_3^{(2)}) = 3\times 2$$

$$(C_3^{(1)})_{ij} = \begin{cases} \theta_1\,; & i=j=1 \\ 0\,; & \text{elsewhere} \end{cases}, (C_3^{(2)})_{ij} = \begin{cases} \mu_1\,; & i=j=1 \\ 0\,; & \text{elsewhere} \end{cases}$$

$$A_{00}^{(3)} = \begin{pmatrix} -\lambda & 0 & & & \\ M & E_3^{(0)} & & & \\ & E_3^{(2)} & E_3^{(1)} & & \\ & & E_3^{(2)} & E_3^{(1)} & \\ & & & \ddots & \ddots \\ & & & & \end{pmatrix}, M = \begin{pmatrix} \begin{bmatrix} \mu_2 \\ \mu_3 \end{bmatrix} \\ \mathbf{0} \\ \mathbf{0} \\ \vdots \end{pmatrix}$$

$$E_3^{(1)} = I_\infty \otimes D_{31},\ \dim(D_{31}) = 2,\ (D_{31})_{ij} = \begin{cases} -(\lambda+\mu_2+\theta_2)\,; & i=j=1 \\ -(\lambda+\mu_3)\,; & i=j=2 \end{cases}$$

$$E_3^{(2)} = E_3^{(21)} + E_3^{(22)}$$

$$E_3^{(21)} = I_\infty \otimes D_3^{(2)}, \dim(D_3^{(2)}) = 2,\ \left(D_3^{(2)}\right)_{ij} = \begin{cases} \mu_{i+1}\,; & j=1, i=1,2 \\ 0\,; & \text{elsewhere} \end{cases}$$

$$E_3^{(22)} = \begin{pmatrix} 0 & g_3^{(2)} & & \\ & & g_3^{(2)} & \\ & & & \ddots \\ & & & \end{pmatrix}, \quad \begin{array}{l} \dim(g_3^{(2)}) = 2 \\ \left(g_3^{(2)}\right)_{ij} = \begin{cases} \theta_2\,; & j=i=1 \\ 0\,; & \text{elsewhere} \end{cases} \end{array}$$

$$E_3^{(0)} = \begin{pmatrix} J_3^{(1)} & & & \\ J_3^{(11)} & J_3^{(1)} & & \\ & J_3^{(11)} & J_3^{(1)} & \\ & & \ddots & \ddots \end{pmatrix}$$

$$\dim(J_3^{(1)}) = \dim(J_3^{(11)}) = 2$$

$$\left(J_3^{(1)}\right)_{ij} = \begin{cases} -(\lambda+\mu_2+\theta_2) & ; \quad i=j=1 \\ -(\lambda+\mu_3) & ; \quad i=j=2 \\ \theta_2 & ; \quad i=1, j=2 \\ 0 & ; \quad \text{elsewhere} \end{cases}, \left(J_3^{(11)}\right)_{ij} = \begin{cases} \mu_{i+1} ; & i=1,2 ; j=2 \\ 0 \quad ; & \text{elsewhere} \end{cases}$$

Theorem 4.1. *Let $\boldsymbol{x} = [\boldsymbol{x}_0, \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots]$ denote the invariant probability vector for the QBD process with infinite number of sub levels, where $\boldsymbol{x}_i$ is the probability vector of infinite dimension corresponding to level i . Then the solution for $\boldsymbol{x}$ possesses a matrix geometric structure*

$$\boldsymbol{x}_i = \boldsymbol{x}_{i-1} R, \ i > 1. \tag{4.1}$$

where the rate matrix R is the minimal non negative solution to

$$R^2 A_2 + R A_1 + A_0 = 0. \tag{4.2}$$

Theorem 4.2. *The rate matrix R in the above theorem possesses a block upper triangular structure given by*

$$R = \begin{pmatrix} R_0 & R_1 & R_2 & \cdots \\ & R_0 & R_1 & \cdots \\ & & R_0 & \cdots \\ & & & \ddots \end{pmatrix}, \textit{where } R_j = \begin{pmatrix} R_{j0} & R_{j1} & R_{j2} & \cdots \\ & R_{j0} & R_{j1} & \cdots \\ & & R_{j0} & \cdots \\ & & & \ddots \end{pmatrix}; j \geq 0,$$

in which R_{ji} are matrices of the form $R_{ji} = \begin{pmatrix} a_{ji} & 0 & 0 \\ b_{ji} & c_{ji} & 0 \\ d_{ji} & 0 & f_{ji} \end{pmatrix}; i \geq 0,$ whose entries are as follows: Let $x = \lambda + \mu_1 + \theta_1,\ y = \lambda + \mu_2 + \theta_2,$ and $z = \lambda + \mu_3$. Then,

$$a_{00} = \frac{x - \sqrt{x^2 - 4\mu_1\lambda}}{2\mu_1}, \ b_{00} = \frac{2\mu_2\lambda}{xy + y\sqrt{x^2 - 4\mu_1\lambda} - 2\mu_1\lambda}, \ c_{00} = \frac{\lambda}{y},$$

$$d_{00} = \frac{2\mu_3\lambda}{xz + z\sqrt{x^2 - 4\mu_1\lambda} - 2\mu_1\lambda}, \ f_{00} = \frac{\lambda}{z}, a_{0k} = 0; \ k \geq 1,$$

$$b_{01} = \frac{\theta_2 c_{00}}{x - (a_{00} + c_{00})\mu_1}, \ b_{0k} = 0; k \geq 2, c_{0k} = d_{0k} = f_{0k} = 0; \ k \geq 1,$$

$$a_{l0} = \frac{\theta_1 \sum\limits_{k=0}^{l-1} a_{k0} a_{(l-1-k)0} + \mu_1 \sum\limits_{k=1}^{l-1} a_{k0} a_{(l-k)0}}{x - 2a_{00}\mu_1}; \ l \geq 1,$$

$$b_{l0} = \frac{\mu_1 \sum\limits_{i=1}^{l} a_{i0} b_{(l-i)0} + \theta_1 \left[\sum\limits_{j=0}^{l-1} a_{j0} b_{(l-1-j)0} + b_{(l-1)0} c_{00} \right]}{x - (a_{00} + c_{00})\mu_1}; \ l \geq 1,$$

$$d_{l0} = \frac{\mu_1 \sum\limits_{i=1}^{l} a_{i0} d_{(l-i)0} + \theta_1 \left[\sum\limits_{j=0}^{l-1} a_{j0} d_{(l-1-j)0} + d_{(l-1)0} f_{00} \right]}{x - (a_{00} + f_{00})\mu_1}; \; l \geq 1,$$

$$c_{l0} = f_{l0} = 0; \; l \geq 1,$$

$$a_{mn} = c_{mn} = d_{mn} = f_{mn} = 0; \; m, n \geq 1,$$

$$b_{m1} = \frac{\mu_1 \sum\limits_{i=1}^{l} a_{i0} b_{(m-i)1} + \theta_1 \left[\sum\limits_{j=0}^{m-1} a_{j0} b_{(m-1-j)1} + b_{(m-1)1} c_{00} \right]}{x - (a_{00} + c_{00})\mu_1}; \; m \geq 1,$$

$$b_{mn} = 0; \; m \geq 1, n \geq 2.$$

Proof: Expansion of equation (4.2) gives the following system of equations:

$$R_{00}^2 M_3 + R_{00} L_3 + \lambda I_3 = 0 \tag{4.3}$$

$$\sum_{j=0}^{m} R_{0j} R_{0,m-j} M_3 + R_{0,m-1} U_3^{(2)} + R_{0m} L_3 = 0 \tag{4.4}$$

$$\sum_{k=0}^{l-1} R_{k0} R_{l-1-k,0} N_3 + \sum_{k=0}^{l} R_{k0} R_{l-k,0} M_3 + R_{l0} M_1 = 0 \tag{4.5}$$

(4.6)

$$\sum_{k=0}^{l-1} \sum_{j=0}^{m} R_{kj} R_{l-1-k,m-j} N_3 + \sum_{k=0}^{l} \sum_{j=0}^{m} R_{kj} R_{l-k,m-j} M_3 + R_{l,m-1} U_3^{(2)} + R_{lm} L_3 = 0. \tag{4.7}$$

Solving the system required result is obtained.

4.1. **Joint and Marginal Probabilities.** Let $\boldsymbol{x}_{ijk}$ be the probability of i high priority customers in the system, j customers waiting in the $\mathcal{P}_2$ queue and k customers waiting in the $\mathcal{P}_3$ queue.

Then the marginal probability of i number of $\mathcal{P}_1$ customers is

$$\boldsymbol{x}_{i..} = \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \boldsymbol{x}_{ijk}.$$

We have, from theorem (4.1) $\boldsymbol{x}_i = \boldsymbol{x}_{i-1} R$ and proceeding as in the section 3,

$$\boldsymbol{x}_{ijk} = \sum_{l=0}^{j} \sum_{m=0}^{k} \boldsymbol{x}_{i-1,lm} R_{j-l,k-m} \text{ for} j, k \geq 0.$$

To know the type of customer under service, we expand the above equation to get the recursive formulas,

$$\begin{aligned}
\boldsymbol{x}_{ijk}(1) &= \sum_{l=0}^{j} a_{j-l,0} \boldsymbol{x}_{i-1,lk}(1) + \sum_{l=0}^{j} \sum_{m=0}^{l} b_{j-l,l-m} \boldsymbol{x}_{i-1,lm}(2) + \sum_{l=0}^{j} d_{j-l,0} \boldsymbol{x}_{i-1,lk}(3), \\
\boldsymbol{x}_{ijk}(2) &= c_{00} \boldsymbol{x}_{i-1,jk}(2), \\
\boldsymbol{x}_{ijk}(3) &= f_{00} \boldsymbol{x}_{i-1,jk}(3); \; j, k \geq 0, i \geq 1.
\end{aligned}$$

4.2. **High Priority Marginal Distribution.** Marginal distribution of high priority customers in the system is

$$\begin{aligned}\boldsymbol{x}_{i..} &= \sum_{j=0}^{\infty}\sum_{k=0}^{\infty} x_{ijk}\\ &= x_{i-1,..}R_{+}\end{aligned}$$

where

$$\begin{aligned}\boldsymbol{R}_{+} &= \sum_{j=0}^{\infty} R_j\\ &= \sum_{j=0}^{\infty}\sum_{k=0}^{\infty} R_{jk}.\end{aligned}$$

Expanding the above equation in lowest phases gives,

$$x_{i..} = [x_{i..}(1),\ x_{i..}(2),\ x_{i..}(3)] = [(1-\rho),\ x_0(2),\ x_0(3)]\,\boldsymbol{R}_{+}^{i}, \quad i \geq 1.$$

5. Case of $N+1$ priorities, non-preemptive:

Model Description: Here also we consider a single server infinite capacity queuing system in which customers from outside arrive according to a Poisson process with rate λ and form a queue if server is busy. Service times are exponentially distributed with parameter μ_1. Customers in primary queue interrupt service according to a Poisson process of rate θ_1, in which case he has to go to a lower priority queue. Else, he completes service and leaves the system forever. Lower priority customers are taken for service according to head of the line priority whenever the queue of external customers is found to be empty at a service completion epoch. The service of such customers is according to a non-preemptive service discipline. A customer from this low priority queue may interrupt his service according to a Poisson process of rate θ_2 up on which he has to go to a third waiting line (of infinite capacity) and wait for his turn for service. The service time of customers in the i_{th} queue are independent and identically distributed exponential random variables with parameter μ_i. Customers in the i^{th} priority queue also interrupt their service according to a Poisson process with rate θ_i or else completes service with service time exponentially distributed with parameter μ_i. A maximum of N service interruptions is allowed for any customer so that $i = 2, 3, \ldots, N$. Thus there are $N+1$ queues, the first one constituted solely by external (primary) customers and the remaining queues are generated by customers from the just preceding higher priority queue. Thus N dependent queues and one independent stream of customers served by a single server, form our system. At the service completion epoch of a low priority customer, the server checks whether there is any higher priority customer in the system. If there is one in the highest priority, he takes the head in that queue; else takes the one, if any, from the second queue and so on. From the $(N+1)^{th}$ queue, a customer in service leaves on completion of service (following an exponential distribution with parameter μ_{N+1}) or interrupts his service according to a Poisson process of rate θ_{N+1}. In the latter case the customer leaves the system paying a heavy penalty.

The infinitesimal generator is

$$\widehat{Q} = \begin{pmatrix} A_{00}^{(n)} & A_{01}^{(n)} & & & \\ A_{10}^{(n)} & A_{1}^{(n)} & A_{0}^{(n)} & & \\ & A_{2}^{(n)} & A_{1}^{(n)} & A_{0}^{(n)} & \\ & & \ddots & \ddots & \end{pmatrix}$$

where,

$$A_0^{(n)} = \lambda I_\infty, A_1^{(n)} = I_\infty \otimes H_n$$

$$H_n = \begin{pmatrix} L_n & U_n^{(n-1)} & 0 & ... & U_n^{(n-2)} & 0 & ... & U_n^{(3)} & 0 & ... & U_n^{(2)} & 0 & ... & ... \\ 0 & L_n & U_n^{(n-1)} & 0 & ... & U_n^{(n-2)} & 0 & ... & U_n^{(3)} & 0 & ... & U_n^{(2)} & 0 & ... \\ & & & \ddots & \ddots & & & & \ddots & & & & & \\ & & & & \ddots & \ddots & & & & \ddots & & & & \end{pmatrix}$$

$\dim(Ln) = n,\ n \geq 3,\ \dim(Un^{(k)}) = n,\ k = 2, 3, ..., n-1$

$$(L_n)_{ij} = \begin{cases} -(\lambda + \theta_i + \mu_i) & j = i = 1, 2, ..., n-1. \\ -(\lambda + \mu_n) & j = i = n \\ \mu_i & j = 1, i = 2, 3, ..., n \\ 0 & \text{otherwise} \end{cases}, (Un^{(k)})_{ij} = \begin{cases} \theta_k & j = 1, i = k \\ 0 & \text{otherwise} \end{cases}$$

$$\dim(Mn) = \dim(Nn) = n$$

$$A_2^{(n)} = \begin{pmatrix} I_\infty \otimes M_n & I_\infty \otimes N_n & & \\ & I_\infty \otimes M_n & I_\infty \otimes N_n & \\ & & \ddots & \ddots \\ & & & \ddots \\ & & & \end{pmatrix}, \quad (M_n)\,ij = \begin{cases} \mu_1 & ; j = i = 1 \\ 0 & ; \text{otherwise} \end{cases}$$

$$(N_n)\,ij = \begin{cases} \theta_1 & ; j = i = 1 \\ 0 & ; \text{otherwise} \end{cases}$$

$$A_{10}^{(n)} = \begin{pmatrix} C_n^* & C_n^{(0)} + I_\infty \otimes C_n^{(1)} & & & \\ & I_\infty \otimes C_n^{(2)} & I_\infty \otimes C_n^{(1)} & & \\ & & I_\infty \otimes C_n^{(2)} & I_\infty \otimes C_n^{(1)} & \\ & & & \ddots & \ddots \\ & & & & \ddots \\ & & & & \end{pmatrix}$$

$$C_n^* = \begin{pmatrix} \mu_1 \\ 0 \\ 0 \\ \vdots \\ \end{pmatrix}, \ C_n^{(0)} = \begin{pmatrix} \mathbf{0} \\ I_\infty \otimes B_1 \\ I_\infty \otimes B_2 \\ \cdots \\ \cdots \\ I_\infty \otimes B_{n-2} \end{pmatrix}, \quad \begin{array}{l} \dim(B_k) = n \times (n-1), 1 \leq k \leq (n-k), \\ (B_k)_{ij} = \begin{cases} \mu_1 \ ; & i = 1, j = n-k \\ 0 \ ; & \text{elsewhere} \end{cases} \end{array}$$

and $\mathbf{0} = [\ 0 \quad 0 \quad \cdots \quad]$

$\dim(C_n^{(1)}) = \dim(C_n^{(2)}) = n \times (n-1)$

$$(C_n^{(1)})_{ij} = \begin{cases} \theta_1\,; & i=j=1 \\ 0\,; & \text{elsewhere} \end{cases} \quad (C_n^{(2)})_{ij} = \begin{cases} \mu_1\,; & i=j=1 \\ 0\,; & \text{elsewhere} \end{cases}$$

$$A_{01}^{(n)} = \begin{pmatrix} K_n^{(0)} & & & & \\ I_\infty \otimes K_n & & & & \\ & I_\infty \otimes K_n & & & \\ & & I_\infty \otimes K_n & & \\ & & & \ddots & \ddots \end{pmatrix},$$

$K_n^{(0)} = [\ \lambda \quad 0 \quad 0 \quad ...\ ], dim(K_n) = (n-1) \times n,$

$(K_n)_{ij} = \begin{cases} \lambda\,; & j = i+1, i = 1, 2, ..., (n-1). \\ 0\,; & elsewhere. \end{cases}$

$$A_{00}^{(n)} = \begin{pmatrix} -\lambda & & & & \\ M & E_n^{(0)} & & & \\ & E_n^{(2)} & E_n^{(1)} & & \\ & & E_n^{(2)} & E_n^{(1)} & \\ & & & \ddots & \ddots \\ & & & & \end{pmatrix}, M = \begin{pmatrix} \begin{bmatrix} \mu_2 \\ \mu_3 \\ \vdots \\ \mu_n \end{bmatrix} \\ \mathbf{0} \\ \mathbf{0} \\ \vdots \end{pmatrix}$$

$$E_n^{(1)} = I_\infty \otimes D_{n1}, \dim(D_{n1}) = n-1, (D_{n1})_{ij} = \begin{cases} -(\lambda + \mu_{i+1} + \theta_{i+1}) & 1 \le i, j \le n-2 \\ -(\lambda + \mu_{i+1}) & i = j = n-1 \end{cases}$$

$$E_n^{(2)} = E_n^{(21)} + E_n^{(22)}$$

$$E_n^{(21)} = I_\infty \otimes D_n^{(n-1)}, \dim(D_n^{(n-1)}) = n-1,\ \left(D_n^{(n-1)}\right)_{ij} = \begin{cases} \mu_{i+1} & j = 1, 1 \le i \le n-1 \\ 0 & \text{elsewhere} \end{cases}$$

$$E_n^{(22)} = \begin{pmatrix} 0 & G_n^{(n-1)} & ... & & G_n^{(n-2)} & ... & & G_n^{(2)} & ... & \\ & & G_n^{(n-1)} & ... & & G_n^{(n-2)} & ... & & G_n^{(2)} & ... \\ & & & \ddots & & & \ddots & & & \ddots \\ & & & & & & & & & \end{pmatrix}$$

$$\dim(G_n^{(k)}) = (n-1); (G_n^{(k)})_{ij} = \begin{cases} \theta_k\ ; & i=k-1, j=1 \\ 0\ ; & elsewhere, k=2,3,...,n-1 \end{cases}$$

$$E_n^{(0)} = \begin{pmatrix} J_n^{(1)} & & & & & & & & & \\ J_n^{(11)} & J_n^{(1)} & & & & & & & & \\ & J_n^{(11)} & J_n^{(1)} & & & & & & & \\ \vdots & & \ddots & \ddots & & & & & & \\ J_n^{(21)} & J_n^{(22)} & \cdots & & J_n^{(2)} & & & & & \\ & J_n^{(21)} & J_n^{(22)} & \cdots & & J_n^{(2)} & & & & \\ \vdots & & \ddots & \ddots & & & \ddots & & & \\ J_n^{(31)} & J_n^{(32)} & \cdots & & J_n^{(33)} & \cdots & J_n^{(3)} & & & \\ & & \ddots & & & \ddots & & \ddots & & \\ J_n^{((n-2)1)} & J_n^{((n-2)2)} & \cdots & & J_n^{((n-2)3)} & \cdots & & & J_n^{((n-2))} & \\ \vdots & & \ddots & \ddots & & & & & & \ddots \end{pmatrix}$$

$$\dim(J_n^{(k)}) = \dim(J_n^{(km)}) = n-1\ ; \begin{cases} k=1,2,...,,(n-2) \\ m=1,2,...,,k \end{cases}$$

$$\left(J_n^{(k)}\right)_{ij} = \begin{cases} -\left(\lambda + \mu_{i+1} + \theta_{i+1}\right)\ ; & i=j=1,2,...,(n-2) \\ -\left(\lambda + \mu_{i+1}\right)\ ; & i=j=(n-1) \\ \theta_{i+1}\ ; & j=i+1, i=1,2,...,(n-k+1) \\ 0\ ; & elsewhere \end{cases}$$

$$\left(J_n^{(k1)}\right)_{ij} = \begin{cases} \mu_{i+1}\ ; & i=1,2,...,(n-1)\ ; j=n-k \\ 0\ ; & elsewhere \end{cases}$$

$$\left(J_n^{(km)}\right)_{ij} = \begin{cases} \theta_{n-m+1}\ ; & i=n-m,\ ; j=n-k \\ 0\ ; & elsewhere,\ : m=2,3,...,k, \end{cases}$$

The stability of the system is given to be

$$\frac{\lambda}{\mu_1+\theta_1}\left[1 + \sum_{i=1}^{N-1}\prod_{j=1}^{i}\frac{\theta_j}{\mu_{j+1}+\theta_{j+1}} + \frac{\theta_N}{\mu_{N+1}}\prod_{j=1}^{N-1}\frac{\theta_j}{\mu_{j+1}+\theta_{j+1}}\right] < 1.$$

CONCLUSION

In this paper we considered a highly dependent priority queueing system where low priority customers join the queue from immediately preceding waiting lines. We assumed that all the underlying distributions are exponential. Analytical expressions for system state probabilities were computed. The case where external customers joining low priority queues, depending on their priority, is being analyzed. Further the presented model is extended to the case where service times are distinct phase type distributions, depending on the priority class.

ACKNOWLEDGMENTS: The first author's research is supported by KSCSTE, Govt. of Kerala, India (No. 001/KESS/2013/CSTE). Second author is indebted to the University Grants Commission, Government of India, for his fellowship under the Faculty Development Programme (Grant No. F.FIP/12thPlan/KLMG003TF05).

REFERENCES

[1] *Cobham, A.* (1954): Priority Assignment in Waiting Line Problems. Operations Research 2, 70-76.

[2] *Dudin, A. N., Jacob, V. and Krishnamoorthy, A.* (2013) : A multi-server queueing system with service interruption, partial protection and repetition of service. Annals of Operations Research, DOI 10.1007/s10479- 013-1318-3, Springer.

[3] *Jaiswal, N. K.* (1961) : Preemptive resume priority queue. Operations Research 9, 732-742.

[4] *Jaiswal, N. K.* (1962) : Time dependent solution of the head of the line priority queue. Jr. Roy Statistic Soc. B. 24, 73-90.

[5] *Jaiswal, N. K.* (1968) : Priority queues, Academic Press, New York and London.

[6] *Krishnamoorthy, A. and Jacob, V.* (2012) : Analysis of Customer Induced Interruption in a multi server system. Neural,Parallel and Scientific Computations, 20, 153–172. Dynamic publishers.

[7] *Miller, D.R.* (1981) : Computation of Steady-State Probabilities for M/M/1 Priority queues, Operations Research.

[8] *Neuts, M.F.* (1981) : Matrix-Geometric Solutions in Stochastic Models, Johns Hopkins University Press, Baltimore.

[9] *Henk C. Tijms* (1998) : Stochastic Models-An Algoritheoremic Approach, John Wiley & Sons.

[10] *Sapna Isotupa K. P., David A. Stanford* (2002) : An Infinite-Phase Quasi-Birth-And-Death Model for the Non-preemptive Priority M/Ph/1 Queue, Staochastic Models, 18:3, 387-424.

[11] *Takagi, H.* (1991): Queueing Analysis. Volume 1: Vacations and Priority Systems; North-Holland: Amsterdam.

[12] *Takagi, H. and Kodera, Y.* (1996) : Analysis of preemptive loss priority queues with preemption distance, Queueing Systems 22, 367-381

[13] *Jacob, V., Chakravarthy, S. R. and Krishnamoorthy, A.* (2012) : On a Customer Induced Interruption in a service system. Stochastic Analysis and Applications, 30 : 6, 949–962, DOI: 10.1080/07362994.2012.704845. Taylor & Francis

[14] *White, H. and Christie, L. S.* (1958) : Queueing with Preemptive Priorities or with Breakdown. Operations Research 6, 79-96.