# Multi-class, multi-server queues with non-preemptive priorities

Andrei Sleptchenko.

EURANDOM, Eindhoven University of Technology,

P.O. Box 513, 5600MB Eindhoven, The Netherlands.

e-mail:Sleptchenko@eurandom.tue.nl

**Abstract**

In this paper we consider a multi-class, multi-server queueing system with non-preemptive priorities. We distinguish two groups of priority classes that consist of multiple items, each having their own arrival and service rate. We assume Poisson arrival processes and exponentially distributed service times. We derive an approximation method to estimate the steady state probabilities with an approximation error that can be made as small as desired at the expense of some more numerical matrix iterations. Based on these probabilities, we can derive approximations for a wide range of relevant performance characteristics, such as the expected postponement time for each item class and the first and second moment of the number of items of a certain type in the system.

*Keywords*: multi-server, multi-class queue; non-preemptive priority

## 1 Introduction.

This paper continues our studies on multi-class multi-server (MCMS) queues. In previous papers we have studied MCMS queues without priorities [9], with preemptive priority and two priority groups [10] and with two priority groups and outsourcing of high priority items which can not find a free server (preemptive and non-preemptive rules) [11]. In this paper we study *MCMS queues with two priority groups and non-preemptive priority.*

Multi-server priority queueing systems arise normally in the performance analysis of multi-server communication networks with differentiated services.

There is quite some literature on single server priority queueing systems. However, multi-server priority queueing systems have received much less attention. Such models have been studied by Mitrani et al. [8], Gail et al. [2, 3], Kao et al. [5, 6], Kella and Yechiali [7] and Wagner [12, 13, 14]. Results are available both if preemption is allowed, so high priority items may interrupt the service of low priority items, or not. The non-preemptive queues are analyzed most extensively by Wagner [13], who considers multi-server non-preemptive priority systems with a Markovian arrival process, service times having phase type distributions and both finite or infinite queueing space.

To analyze multi-class, multi-server queues with two priority groups each containing several item classes, we proceed as follows. First we construct the equilibrium state equations (section 2). In section 3 we solve the equilibrium equations for the states with high priority items in the queue. We show how to deal with the remaining equations (no high priority items in the queue) in section 4. For these equations, we can only approximate the system state probabilities. Using the (approximate) state probabilities, we can derive various system performance characteristics: expected waiting times per type, expected queue length per type and even correlations between types, expected postponement time per type. In section 5, we show as an example how to derive the first two moments of the number of items in the system for each type and the expected postponement time per type.

# 2 The Model.

## 2.1 Definitions and notation.

We consider a multi-class system with $k$ servers and unlimited waiting room. As mentioned before, customers are processed according to a non-preemptive priority rule; high priority (hp) items form a queue if there is no server available upon arrival. We denote the number of item classes with high (low) priority by $N^h$ ($N^l$). High priority jobs from subclass $i$ arrive according to a Poisson process with rate $\lambda_i^h$ and low priority jobs from subclass $j$ arrive with rate $\lambda_j^l$. The service times of the subclasses are exponentially distributed with rates $\mu_i^h$ and $\mu_j^l$ for high and low priority item classes, respectively. All servers are equal, and if multiple servers are available to process a job, each available server has an equal chance to get this job. We use $\rho_i^h$, $\rho_j^l$ to denote the utilization rates of high and low priority item classes in the system. We denote the total number of high priority items in the queue by $n$ and the number of items in high priority type $i$ by $n_i$. For the low priority items in the queue, we will use the notation $m$ and $m_j$ respectively.

We characterize the system state by four vectors of dimensions $N^h$ and $N^l$, where the components of each vector refer to the (high and low priority) subclasses:

$\overline{s}^h$ and $\overline{s}^l$ – vectors containing the number of high and low priority items in service per item class.

$\overline{w}^h$ and $\overline{w}^l$ – vectors containing the number of high and low priority items in the queue waiting for first service per item class.

Then the systems state probabilities are denoted by $P_{n,m}\left(\overline{w}^h, \overline{s}^h, \overline{w}^l, \overline{s}^l\right)$.

Other general notations used throughout the paper are:

$\Lambda^h$, $\Lambda^l$, $\mu^h$, $\mu^l$ – sums of the arrival rates and service rates for each class, i.e. $\Lambda^h = \sum_{i=1}^{N^h} \lambda_i^h$, $\Lambda^l = \sum_{i=1}^{N^l} \lambda_i^l$ and $\mu^h = \Lambda^h \Big/ \sum_{i=1}^{N^h} \frac{\lambda_i^h}{\mu_i^h} = \frac{\Lambda^h}{k\rho^h}$, $\mu^l = \Lambda^l \Big/ \sum_{i=1}^{N^l} \frac{\lambda_i^l}{\mu_i^l} = \frac{\Lambda^l}{k\rho^l}$, where general utilization rates for each class are $\rho^h = \frac{\Lambda^h}{k\mu^h}$, $\rho^l = \frac{\Lambda^l}{k\mu^l}$, and the total utilization rate is $\rho = \rho^h + \rho^l$.

$a_i^h$, $a_i^l$ – fractions of arrival rates, i.e. $a_i^h = \frac{\lambda_i^h}{\Lambda^h}$, $a_i^l = \frac{\lambda_i^l}{\Lambda^l}$.

$\delta_i^h$, $\delta_i^l$ – perturbations of service rates, i.e. $\left(1 + \delta_i^h\right) = \frac{\mu_i^h}{\mu^h}$, $\left(1 + \delta_i^l\right) = \frac{\mu_i^l}{\mu^h}$, where $\mu^h$ is average service rate of high priority items $\mu^h = \frac{\Lambda^h}{k\rho^h}$.

$\gamma$ – ratio between service rates of high and low priority items $\gamma = \frac{\mu^l}{\mu^h}$

$\overline{\mu}\left(\overline{s}^h, \overline{s}^l\right)$ – sum of service rates of all items in service, i.e. $\overline{\mu}\left(\overline{s}^h, \overline{s}^l\right) = \sum_{i=1}^{N^h} s_i^h \mu_i^h + \sum_{i=1}^{N^l} s_i^l \mu_i^l$

$\overline{\delta}\left(\overline{s}^h, \overline{s}^l\right)$ – sum of perturbations of service rate of all items in service, i.e. $\overline{\delta}\left(\overline{s}^h, \overline{s}^l\right) = \frac{1}{k}\left(\sum_{i=1}^{N^h} s_i^h \delta_i^h + \sum_{i=1}^{N^l} s_i^l \delta_i^l\right)$

$x_i$ – the $i^{th}$ component of any vector $\overline{x}$.

$\overline{e}_i^h$ ($\overline{e}_i^l$) – a vector of dimension $N^h$ ($N^l$) with component $i$ equal to 1 and all other components equal to 0; this vector is used to indicate the changes in vectors $\overline{w}^h$ and $\overline{s}^h$ ($\overline{w}^l$, $\overline{r}^l$ and $\overline{s}^l$) during transitions from state to state.

$e_{ij}^h$ ($e_{ij}^l$) – denotes the $j^{th}$ component of the vector $e_i^h$ ($e_i^l$), so $e_{ij}^h$ ($e_{ij}^l$) $= 1$ if $i = j$ and 0 otherwise.

$|\overline{x}|$ – denotes the sum of all components of any vector $\overline{x}$.

We will introduce the remaining notation later on.

## 2.2 Stationary state equations.

We divide the equilibrium equations into three groups:

*I.* there is at least one high priority item in the queue ($n > 0$, $m \geq 0$),

*II.* all servers are busy, but there is no high priority item and at least one low priority item in the queue ($n = 0$, $m > 0$),

*III.* there is no queue ($n = m = 0$),

All these subspaces have different equilibrium equations. Besides, we have to consider the equations for the two boundaries between the regions separately.

In **area I** ($n > 0$, $m \geq 0$), we have states with all servers occupied and high priority items in the queue.

The equilibrium equations in this area are :

$$\begin{aligned}
&\left(\Lambda^h + \Lambda^l + \overline{\mu}\left(\overline{s}^h, \overline{s}^l\right)\right) P_{n,m}\left(\overline{w}^h, \overline{s}^h, \overline{w}^l, \overline{s}^l\right) \\
&= \sum_{i=1}^{N^h} \lambda_i^h P_{n-1,m}\left(\overline{w}^h - \overline{e}_i^h, \overline{s}^h, \overline{w}^l, \overline{s}^l\right) + \sum_{i=1}^{N^l} \lambda_i^l P_{n,m-1}\left(\overline{w}^h, \overline{s}^h, \overline{w}^l - \overline{e}_i^l, \overline{s}^l\right) \qquad (1) \\
&+ \sum_{i=1}^{N^h}\sum_{j=1}^{N^h} \frac{w_j^h + 1}{\left|\overline{w}^h\right| + 1}\left(s_i^h + 1 - e_{ij}^h\right) \mu_i^h P_{n+1,m}\left(\overline{w}^h + \overline{e}_j^h, \overline{s}^h + \overline{e}_i^h - \overline{e}_j^h, \overline{w}^l, \overline{s}^l\right) \\
&+ \sum_{i=1}^{N^l}\sum_{j=1}^{N^h} \frac{w_j^h + 1}{\left|\overline{w}^h\right| + 1}\left(s_i^l + 1\right) \mu_i^l P_{n+1,m}\left(\overline{w}^h + \overline{e}_j^h, \overline{s}^h - \overline{e}_j^h, \overline{w}^l, \overline{s}^l + \overline{e}_i^l\right)
\end{aligned}$$

where $\frac{w_j^h+1}{|\overline{w}^h|+1}$ characterizes the probability that the high priority item with class $j$ is in front of the queue.

In **area II** ($n=0$,$m>0$) all servers are busy, low priority items are in service and no high priority items are in the queue (i.e. $\overline{w}^h=0$). Then, the equilibrium equations are:

$$\begin{aligned}
&\left(\Lambda^h+\Lambda^l+\overline{\mu}\left(\overline{s}^h,\overline{s}^l\right)\right)P_{0,m}\left(\overline{0},\overline{s}^h,\overline{w}^l,\overline{s}^l\right)=\sum_{i=1}^{N^l}\lambda_i^l P_{0,m-1}\left(\overline{0},\overline{s}^h,\overline{w}^l-\overline{e}_i^l,\overline{s}^l\right)\\
&+\sum_{i=1}^{N^h}\sum_{j=1}^{N^l}\frac{w_j^l+1}{\left|\overline{w}^l\right|+1}\left(s_i^h+1\right)\mu_i^h P_{0,m+1}\left(\overline{0},\overline{s}^h+\overline{e}_i^h,\overline{w}^l+\overline{e}_j^l,\overline{s}^l-\overline{e}_j^l\right)\\
&+\sum_{i=1}^{N^l}\sum_{j=1}^{N^l}\frac{w_j^l+1}{\left|\overline{w}^l\right|+1}\left(s_i^l+1-e_{ij}^l\right)\mu_i^l P_{0,m+1}\left(\overline{0},\overline{s}^h,\overline{w}^l+\overline{e}_j^l,\overline{s}^l+\overline{e}_i^l-\overline{e}_j^l\right)\\
&+\sum_{i=1}^{N^h}\sum_{j=1}^{N^h}\left(s_i^h+1-e_{ij}^h\right)\mu_i^h P_{1,m}\left(\overline{e}_j^h,\overline{s}^h+\overline{e}_i^h-\overline{e}_j^h,\overline{w}^l,\overline{s}^l\right)\\
&+\sum_{i=1}^{N^l}\sum_{j=1}^{N^h}\left(s_i^l+1\right)\mu_i^l P_{1,m}\left(\overline{e}_j^h,\overline{s}^h-\overline{e}_j^h,\overline{w}^l,\overline{s}^l+\overline{e}_i^l\right)
\end{aligned}\tag{2}$$

Finally, we write down the equilibrium equations for the states without queue. We denote these states as $P_{-r}\left(\overline{0},\overline{s}^h,\overline{0},\overline{s}^l\right)$ where $r$ is equal to the number of empty servers: $r=k-\left|\overline{s}^h\right|+\left|\overline{s}^l\right|$. Note that $P_{-r}\left(\overline{0},\overline{s}^h,\overline{0},\overline{s}^l\right)\equiv P_{0,0}\left(\overline{0},\overline{s}^h,\overline{0},\overline{s}^l\right)$.

The equilibrium equations on the **border between the areas II and III** ($\left|\overline{s}^h\right|+\left|\overline{s}^l\right|=k$, $m=n=0$) have states from both areas II and III. That is:

$$\begin{aligned}
&\left(\Lambda^h+\Lambda^l+\overline{\mu}\left(\overline{s}^h,\overline{s}^l\right)\right)P_0\left(\overline{0},\overline{s}^h,\overline{0},\overline{s}^l\right)=\\
&\sum_{i=1}^{N^h}\lambda_i^h P_{-1,}\left(\overline{0},\overline{s}^h-\overline{e}_i^h,\overline{0},\overline{s}^l\right)+\sum_{i=1}^{N^l}\lambda_i^l P_{-1}\left(\overline{0},\overline{s}^h,\overline{0},\overline{s}^l-\overline{e}_i^l\right)\\
&+\sum_{i=1}^{N^h}\sum_{j=1}^{N^l}\left(s_i^h+1\right)\mu_i^h P_{0,1}\left(\overline{0},\overline{s}^h+\overline{e}_i^h,\overline{e}_j^l,\overline{s}^l\right)+\sum_{i=1}^{N^l}\sum_{j=1}^{N^l}\left(s_i^l+1\right)\mu_i^l P_{0,1}\left(\overline{0},\overline{s}^h,\overline{e}_j^l,\overline{s}^l+\overline{e}_i^l\right)\\
&+\sum_{i=1}^{N^h}\sum_{j=1}^{N^h}\left(s_i^h+1\right)\mu_i^h P_{1,0}\left(\overline{e}_j^h,\overline{s}^h+\overline{e}_i^h,\overline{0},\overline{s}^l\right)+\sum_{i=1}^{N^l}\sum_{j=1}^{N^h}\left(s_i^l+1\right)\mu_i^h P_{1,0}\left(\overline{e}_i^h,\overline{s}^h,\overline{0},\overline{s}^l+\overline{e}_j^l\right)
\end{aligned}\tag{3}$$

For the inner probability states in **area III** we find:

$$\left(\Lambda^h + \Lambda^l + \overline{\mu}\left(\overline{s}^h, \overline{s}^l\right)\right) P_{-r}\left(\overline{0}, \overline{s}^h, \overline{0}, \overline{s}^l\right) =$$
$$\sum_{i=1}^{N^h} \lambda_i^h P_{-r-1}\left(\overline{0}, \overline{s}^h - \overline{e}_i^h, \overline{0}, \overline{s}^l\right) + \sum_{i=1}^{N^l} \lambda_i^l P_{-r-1}\left(\overline{0}, \overline{s}^h, \overline{0}, \overline{s}^l - \overline{e}_i^l\right)$$
$$+ \sum_{i=1}^{N^h} \left(s_i^h + 1\right) \mu_i^h P_{-r+1}\left(\overline{0}, \overline{s}^h + \overline{e}_i^h, \overline{0}, \overline{s}^l\right) + \sum_{i=1}^{N^l} \left(s_i^l + 1\right) \mu_i^l P_{-r+1}\left(\overline{0}, \overline{s}^h, \overline{0}, \overline{s}^l + \overline{e}_i^l\right)$$

In the next sections, we will show how we can solve these equilibrium equations thereby obtaining the exact system state probabilities. We will address the areas I, II and III in Section 3 and Section 4.

# 3 System states with high priority items in queue ($n > k$).

In this section we focus on area I, so there is at least one high priority item in the queue.

## 3.1 Reducing the set of equations

This reduction can be done using the assumption that, given the total numbers of high and low priority items in the queue, the items are distributed over the item subclasses according to a multinomial distribution (cf. van Harten, Sleptchenko[9] and Sleptchenko et al.[10]):

$$P_{n,m}\left(\overline{w}^h, \overline{s}^h, \overline{w}^l, \overline{s}^l\right) = n! \prod_{i=1}^{N^h} \frac{\left(a_i^h\right)^{w_i^h}}{w_i^h!} m! \prod_{j=1}^{N^l} \frac{\left(a_j^l\right)^{w_j^l}}{w_j^l!} P'_{n,m}\left(\overline{s}^h, \overline{s}^l\right). \tag{4}$$

In this way, we reduce the number of state probabilities to be solved.

We substitute the relation (4) in equation (1) and we divide both sides by $k\mu^h$ to obtain equations that can be transformed in matrix form. That is, all equilibrium equations with the same $n$ and $m$ can be written in a matrix form as:

$$\left(\left(1 + \rho^h + \gamma\rho^l\right) \mathbf{I} + \overline{\delta}^h\right) \mathbf{P}_{n,m} = \rho^h \mathbf{P}_{n-1,m} + \rho^l \mathbf{P}_{n,m-1} + A\mathbf{P}_{n+1,m} \tag{5}$$

where $\mathbf{P}_{n+1,m}$ are vectors containing probabilities $P'_{n,m}\left(\overline{s}^h, \overline{s}^l\right)$ from (4) as components. The dimension of vectors $\mathbf{P}_{n+1,m}$ is equal to the amount of different server states given that all servers are occupied, i.e. this dimension is equal to $d\left(N^h+N^l,k\right)$, with $d\left(x,y\right)=\binom{x+y+1}{y}$. $\overline{\delta}^h$ and $A$ are linear operators on a $d\left(N^h+N^l,k\right)$ – dimensional linear space:

$$\begin{aligned}
\overline{\delta}^h\mathbf{P}_{n+1,m}[\overline{s}^h,\overline{s}^l] &= \left(1+\overline{\delta}\left(\overline{s}^h,\overline{s}^l\right)+\rho^h+\gamma\rho^l\right)\mathbf{P}_{n+1,m}[\overline{s}^h,\overline{s}^l] \\
A\mathbf{P}_{n+1,m}[\overline{s}^h,\overline{s}^l] &= \frac{1}{k}\sum_{i=1}^{N^h}\sum_{j=1}^{N^h}\left(s_i^h+1-e_{ij}^h\right)\left(1+\delta_i^h\right)a_j^h\mathbf{P}_{n+1,m}[\overline{s}^h+\overline{e}_i^h-\overline{e}_j^h,\overline{s}^l] \\
&+\frac{1}{k}\sum_{i=1}^{N^l}\sum_{j=1}^{N^h}\left(s_i^l+1\right)\left(1+\delta_i^l\right)a_j^h\mathbf{P}_{n+1,m}[\overline{s}^h-\overline{e}_j^h,\overline{s}^l+\overline{e}_i^l].
\end{aligned}$$

Solving this matrix equation we can find all state probabilities with $(n>0)$. In the next lemma the structure of the solution of this equation is explained.

**Lemma 1**

*Define the matrix-function* $\mathbf{Z}\left(\xi\right)$ *as solution of*

$$\left(\left(1+\rho^h+\gamma\rho^l\right)\mathbf{I}+\overline{\delta}^h\right)=\rho^h\mathbf{Z}+\rho^l\xi+A\mathbf{Z}^{-1},\quad s.t.\ \left|\sigma\left(\mathbf{Z}\right)\right|>1 \tag{6}$$

*Then*

$$\begin{aligned}
&P_{n,m}\left(\overline{w}^h,\overline{s}^h,\overline{w}^l,\overline{s}^l\right) \\
&=n!\prod_{i=1}^{N^h}\frac{\left(a_i^h\right)^{w_i^h}}{w_i^h!}\left(\prod_{j=1}^{N^l}\frac{\left(a_j^l\right)^{w_j^l}}{w_j^l!}\right)\left(\frac{d}{d\xi}\right)^{\left|\overline{w}^l\right|}\left[\left(\mathbf{Z}^{-1}\left(\xi\right)\right)^{n-k}\mathbf{C}\left(\xi\right)\right]_{\xi=0}\left[\overline{s}^h,\overline{s}^l\right]
\end{aligned} \tag{7}$$

*satisfies all equations for* $m\geq 0$, $n>k$.

*Note that* $\sum_{\overline{w}^h,\overline{w}^l}P\left(\overline{w}^h,\overline{s}^h,\overline{w}^l,\overline{s}^l\right)=\left[\left(\mathbf{Z}^{-1}\left(\xi\right)\right)^{n-k}\mathbf{C}\left(\xi\right)\right]_{\xi=1}\left[\overline{s}^h,\overline{s}^l\right]$. *The notation* $\left[\overline{s}^h,\overline{s}^l\right]$ *in the right hand side refers to the indicated vector component.*

*The proof of this lemma is similar to the one presented in [10], and can be presented on request.*

The probabilities of the system states constructed in this section have a differential form, therefore we will need derivatives of the matrix $\mathbf{Z}$. To find these derivatives is not an easy task since we can not derive an analytical form of the matrix $\mathbf{Z}$, but we can use the equation (6) to find such derivatives iteratively (cf. [10]).

# 4 System states with no high priority items in queue ($n = 0$).

In this section we describe the solution of the equilibrium equations for the states with no high priority items in the queue ($n = 0$).

## 4.1 Reducing the set of equations

As in the previous section, we can reduce the set of equations rewriting:

$$P_{0,m}\left(\overline{0}, \overline{s}^h, \overline{w}^l, \overline{s}^l\right) = m! \prod_{i=1}^{N^l} \frac{\left(a_i^l\right)^{w_i^l}}{w_i^l!} P'_{0,m}\left(\overline{s}^h, \overline{s}^l\right).$$

Note that we can omit two parameters, namely the number of high priority items in the queue per subclass (these are always zero) and the number of low priority items in the queue $m$.

Again, this results in a matrix equation for $n = k$, $m \geq 0$, namely:

$$\mathbf{D}_m \mathbf{P}_{0,m} = \mathbf{E}_m \mathbf{P}_{0,m-1} + \mathbf{B}_m \mathbf{P}_{1,m} + \mathbf{G}_m \mathbf{P}_{0,m+1}$$

where the operators $\mathbf{D}_m$, $\mathbf{E}_m$, $\mathbf{B}_m$ and $\mathbf{G}_m$ on the vectors $\zeta\left[\overline{s}^h, \overline{s}^l\right]$ are respectively defined as:

$$\mathbf{D}_m \zeta\left[\overline{s}^h, \overline{s}^l\right] \stackrel{def}{=} \left(1 + \overline{\delta}\left(\overline{s}^h, \overline{s}^l\right) + \rho^h + \rho^l\right) \zeta\left[\overline{s}^h, \overline{s}^l\right]$$

$$\mathbf{E}_m \zeta\left[\overline{s}^h, \overline{s}^l\right] \stackrel{def}{=} \rho^l \zeta\left[\overline{s}^h, \overline{s}^l\right]$$

$$\mathbf{B}_m \zeta\left[\overline{s}^h, \overline{s}^l\right] \stackrel{def}{=} \begin{cases} \frac{1}{k} \sum_{i=1}^{N^h} \sum_{j=1}^{N^l} a_i^h \left(s_j^h + 1 - e_{ij}^h\right)\left(1 + \delta_j^h\right) \zeta\left[\overline{s}^h - \overline{e}_i^h + \overline{e}_j^l, \overline{s}^l\right] \\ \frac{1}{k} \sum_{i=1}^{N^h} \sum_{j=1}^{N^l} a_i^h \left(s_j^l + 1\right)\left(1 + \delta_j^l\right) \zeta\left[\overline{s}^h - \overline{e}_i^h, \overline{s}^l + \overline{e}_j^l\right] \end{cases}$$

$$\mathbf{G}_m \zeta\left[\overline{s}^h, \overline{s}^l\right] \stackrel{def}{=} \begin{cases} \frac{1}{k} \sum_{i=1}^{N^h} \sum_{j=1}^{N^l} a_j^l \left(s_i^h + 1\right)\left(1 + \delta_i^h\right) \zeta\left[\overline{s}^h + \overline{e}_i^l, \overline{s}^l - \overline{e}_j^l\right] \\ \frac{1}{k} \sum_{i=1}^{N^l} \sum_{j=1}^{N^l} a_j^l \left(s_i^l + 1 - e_{ij}^l\right)\left(1 + \delta_i^l\right) \zeta\left[\overline{s}^h, \overline{s}^l + \overline{e}_i^l - \overline{e}_j^l\right] \end{cases}$$

The matrix $\mathbf{B}_m$ corresponds to probability states with one high priority item in the queue ($n = 1$), which are equal to $\frac{1}{m}\left(\frac{d}{d\xi}\right)^m \left[\mathbf{Z}^{-1}(\xi)\,\mathbf{C}(\xi)\right]$. So, we can write these matrix equations as:

$$\mathbf{D}_m \mathbf{P}_{0,m} = \mathbf{E}_m \mathbf{P}_{0,m-1} + \mathbf{B}_m \frac{1}{m!}\left(\frac{d}{d\xi}\right)^m \left[\mathbf{Z}^{-1}(\xi)\,\mathbf{C}(\xi)\right]_{\xi=0} + \mathbf{G}_m \mathbf{P}_{0,m+1}, \quad m > 0 \tag{8}$$

Next, it is easy to see that the equilibrium equations for the probability states with no items in the queue can be written in matrix form as:

$$\begin{aligned}
&\mathbf{D}_0\mathbf{P}_{0,0} = \mathbf{E}_0\mathbf{P}_{-1} + \mathbf{B}_0\left[\mathbf{Z}^{-1}(\xi)\,\mathbf{C}(\xi)\right]_{\xi=0} + \mathbf{G}_0\mathbf{P}_{0,1}, \quad r=0 \\
&\mathbf{D}_{-r}\mathbf{P}_{-r} = \mathbf{E}_{-r}\mathbf{P}_{-r-1} + \mathbf{G}_{-r}\mathbf{P}_{-r}, \quad 0<r<k \\
&\mathbf{D}_{-k}\mathbf{P}_{-k} = \mathbf{G}_{-k}\mathbf{P}_{-k+1}, \quad r=k
\end{aligned} \tag{9}$$

where the operators $\mathbf{D}_{-r}$, $\mathbf{E}_{-r}$ and $\mathbf{G}_{-r}$ on the vectors $\zeta\left[\overline{s}^h,\overline{s}^l\right]$ are respectively defined as:

$$\mathbf{D}_{-r}\zeta\left[\overline{s}^h,\overline{s}^l\right] \stackrel{def}{=} \left(\frac{k-r}{k} + \overline{\delta}\left(\overline{s}^h,\overline{s}^l\right) + \rho^h + \rho^l\right)\zeta\left[\overline{s}^h,\overline{s}^l\right]$$

$$\mathbf{E}_{-r}\zeta\left[\overline{s}^h,\overline{s}^l\right] \stackrel{def}{=} \begin{cases} \frac{1}{k}\sum\limits_{i=1}^{N^h}\zeta\left[\overline{s}^h-\overline{e}_i^l,\overline{s}^l\right] \\ \frac{1}{k}\sum\limits_{j=1}^{N^l}\zeta\left[\overline{s}^h,\overline{s}^l-\overline{e}_i^l\right] \end{cases}$$

$$\mathbf{G}_{-r}\zeta\left[\overline{s}^h,\overline{s}^l\right] \stackrel{def}{=} \begin{cases} \frac{1}{k}\sum\limits_{i=1}^{N^h}\left(s_i^h+1\right)\left(1+\delta_i^h\right)\zeta\left[\overline{s}^h+\overline{e}_i^l,\overline{s}^l\right] \\ \frac{1}{k}\sum\limits_{i=1}^{N^l}\left(s_i^l+1\right)\left(1+\delta_i^l\right)\zeta\left[\overline{s}^h,\overline{s}^l+\overline{e}_i^l\right] \end{cases}$$

The dimension of the equations in (8) does not depend on $n$ and $m$. As in the case with high priority items in the queue, it can be derived from the number of all combinations of $N^h + N^l$ items on $k$ servers. We find:

$$\dim(\mathcal{L}_{\text{"}|\overline{s}^h|+|\overline{s}^l|>=k\text{"}}) = \binom{N^h+N^l+k-1}{k}. \tag{10}$$

The dimension of the equations (9) depends on $\left|\overline{s}^h\right| + \left|\overline{s}^l\right| = k-r$ and should be equal to the number of all combinations of $N^h + N^l$ items on $k-r$ servers ($r \le k$), i.e.

$$\dim(\mathcal{L}_{\text{"}|\overline{s}^h|+|\overline{s}^l|<k\text{"}}) = \binom{N^h+N^l+k-r-1}{k-r}, \tag{11}$$

### 4.2 States with only low priority items in queue ($m > 0$, $n = 0$).

The probabilities of the system states having only low priority items in queue satisfy the system of linear inhomogeneous difference equations of second order with fixed coefficients (8). However, the inhomogeneous term has a differential form, therefore the standard procedure of solving the inhomogeneous equations (solution of homogeneous + partial solution of inhomogeneous) is difficult to apply. Therefore we will look for the solution in a differential form $\mathbf{P}_{0,m} = \frac{1}{m!}\left(\frac{d}{d\xi}\right)^m v(\xi)_{\xi=0}$. The substitution of this solution into the equation (8) gives:

$$\mathbf{D}_m \frac{1}{m!}\left(\frac{d}{d\xi}\right)^m v(\xi)_{\xi=0} = \mathbf{E}_m \frac{1}{(m-1)!}\left(\frac{d}{d\xi}\right)^{m-1} v(\xi)_{\xi=0} + \mathbf{B}_m \frac{1}{m!}\left(\frac{d}{d\xi}\right)^m \left[\mathbf{Z}^{-1}(\xi)\,\mathbf{C}(\xi)\right]_{\xi=0}$$
$$+\,\mathbf{G}_m \frac{1}{(m+1)!}\left(\frac{d}{d\xi}\right)^{m+1} v(\xi)_{\xi=0}\,, \quad m>0$$

Here we can apply the following equations: $\left(\frac{d}{dx}\right)^t (xf(x))_{x=0} = t\left[\left(\frac{d}{dx}\right)^{t-1} f(x)\right]_{x=0}$ and $\left(\frac{d}{dx}\right)^t \left(x^2 f(x)\right)_{x=0} = t(t-1)\left[\left(\frac{d}{dx}\right)^{t-2} f(x)\right]_{x=0}$ that can be easily proved as shown in the proof of lemma 1 (Appendix 1). These two equations allow us to remove the derivatives from equation (8) and to obtain a new expression of the function $v(\xi)$ for any $t>0$:

$$\frac{1}{(m+1)!}\left(\frac{d}{d\xi}\right)^{m+1} \left[\xi\mathbf{D}v(\xi) - \xi^2\mathbf{E}v(\xi) - \mathbf{G}v(\xi) - \xi\mathbf{B}\mathbf{Z}^{-1}(\xi)\,\mathbf{C}(\xi)\right]_{\xi=0} = 0, \quad m>0. \qquad (12)$$

Assume now that the function $\mathbf{C}(\xi)$ is equal to $v(\xi)$, i.e. $\mathbf{C}(\xi) = v_k(\xi)$.

The right part of equation (12) should be a function which becomes zero for any $m>0$, i.e. a linear function. Hence, we obtain another expression for the vector-function $v(\xi)$, that does not contain derivatives, but that contains unknown vectors $C_1$ and $C_2$:

$$\xi\mathbf{D}v(\xi) - \xi^2\mathbf{E}v(\xi) - \mathbf{G}v(\xi) - \xi\mathbf{B}\mathbf{Z}^{-1}(\xi)\,v(\xi) = C_1\xi + C_2 \qquad (13)$$

or

$$\mathbf{H}(\xi)\,v(\xi) = C_1\xi + C_2 \qquad (14)$$

The constants $C_1$ and $C_2$ can be easily expressed via the probability states $\mathbf{P}_{0,0}$, and $\mathbf{P}_{-1}$ by differentiating and setting the equation (13) to 0. That is (see also Sleptchenko et al. [10]):

$$C_2 = -\mathbf{G}\mathbf{P}_k,$$

$$C_1 = \mathbf{E}_0\mathbf{P}_{-1}.$$

In this way, we have defined a function $v(\xi)$ given the probability vectors $\mathbf{P}_{0,0}$ and $\mathbf{P}_{-1}$ and next all probability vectors $\mathbf{P}_{0,m}$ for $m = 1,2,\ldots$ follow from $\mathbf{P}_{0,0}$ and $\mathbf{P}_{-1}$. However, an essential piece of information has not been used up to now. It is clear that we are looking for decaying solutions $\mathbf{P}_{0,m}$ for $m \longrightarrow \infty$. As a consequence $v(\xi)$ should be analytic on a circle with radius $1+\varepsilon$ for some $\varepsilon > 0$. Due to (13) extra conditions have to be satisfied at points $\xi$ inside this circle where $\mathbf{H}(\xi)$ is singular. It turns out that there are several such points in general. For example, $\xi = 0$ and $\xi = 1$ are points of this type. It is easy to check that in case $\xi = 0$ any vector with 0 entries is in the null space of $\mathbb{G}$. Using the equilibrium property for subsystems with $\overline{s}^h, \overline{s}^l$ fixed and $m$ arbitrary it is not difficult to check that $\mathbf{1}^t = (1, ...., 1)$ is a left eigenvalue of $\mathbb{H}(1)$ for the eigenvalue 0. In the next section we shall show that the decay requirement boils down to a relation between the initial conditions $\mathbf{P}_{0,1}$ and $\mathbf{P}_{0,0}$ of the following type:

$$\mathbf{P}_{0,1} = \mathbf{Q}_0 \mathbf{P}_{0,0}.$$

In the next section it will be shown that this can be done by using a direct method without reference to singular points of $\mathbf{H}(\xi)$. This result is crucial in the case with no queue at all.

### 4.3 Decay of $\mathbf{P}_{0,m}$ for $m \longrightarrow \infty$.

From (13) it follows by differentiating $m \geq 2$ times with respect to $\xi$ that:

$$\binom{m}{0} \mathbf{H}(0) \frac{d^m}{d\xi^m} v(0) + \binom{m}{1} \mathbf{H}'(0) \frac{d^{m-1}}{d\xi^{m-1}} v(0) + \cdots + \binom{m}{m} \frac{d^m}{d\xi^m} \mathbb{H}(0)\, v(0) = 0.$$

We can also write this in the form

$$\sum_{i=0}^{m} h_i \mathbf{P}_{0,m-i} = 0,$$

where $\mathbf{P}_{0,i} = \frac{1}{i!} \frac{d^i}{d\xi^i} v(0)$ and $h_i = \frac{1}{i!} \frac{d^i}{d\xi^i} \mathbb{H}(0)$.

An important consequence of this equation is that

$$\mathbf{P}_{0,1} = \mathbf{Q}_0^m \mathbf{P}_{0,0} + \Omega_0^m \mathbf{P}_{0,m} \tag{15}$$

We will derive this result shortly. First we mention that this has the implication that for decaying solutions $\mathbf{P}_{0,m}$ for $m \longrightarrow \infty$ if $\Omega_0^m$ remains bounded:

$$\mathbf{P}_{0,1} = \mathbb{Q}_0 \mathbf{P}_{0,0} \text{ with } \mathbb{Q}_0 = \lim_{m \longrightarrow \infty} \mathbf{Q}_0^m$$

Let us now show how the matrices $\mathbf{Q}_0^m$ and $\Omega_0^m$ can be determined.

Using backward recursion from $m$ to 1 we can show that for any $m$ and $m^*$ $(m > m^*)$ a relation

$$\Theta_0^{m^*}\mathbf{P}_{0,m} + \sum_{i=1}^{m^*} \Theta_i^{m^*}\mathbf{P}_{0,m^*-i} = 0 \tag{16}$$

is valid. That is:

<u>For $m^* = m - 1$</u> it is clear from the equation for $\mathbf{P}_m$ that we can take $\Theta_i^m = h_i$.

<u>For $m^* < m - 1$</u> we have the equation derived in the previous induction step and also the original equation for $\mathbf{P}_{m^*}$:

$$\begin{aligned} \Theta_0^{m^*+1}\mathbf{P}_{0,m} + \Theta_1^{m^*+1}\mathbf{P}_{0,m^*} + \cdots + \Theta_{m^*+1}^{m^*+1}\mathbf{P}_{0,0} &= 0 \\ h_0\mathbf{P}_{0,m^*} + \cdots + h_{m^*}\mathbf{P}_{0,0} &= 0 \end{aligned}$$

or in other terms

$$\Theta_0^{m^*}\mathbb{P}_{0,m} + \Theta_1^{m^*}\mathbb{P}_{0,m^*-1} + \cdots + \Theta_{m^*}^{m^*}\mathbf{P}_{0,0} = 0$$

where the matrices $\Theta_i^{m^*}$ are equal to

$$\begin{aligned} \Theta_0^{m^*} &= h_0\left(\Theta_1^{m^*+1}\right)^{-1}\Theta_0^{m^*+1} \\ \Theta_i^{m^*} &= \left(h_0\left(\Theta_1^{m^*+1}\right)^{-1}\Theta_{i+1}^{m^*+1} - h_i\right), \quad i = 1, \ldots, m^* \end{aligned}$$

So, we have shown that for any $m$ and $m^*$ $(m > m^*)$ the relation (16) is valid. Next, taking $m^* = 2$ we obtain

$$\Theta_0^2\mathbf{P}_{0,m} + \Theta_1^2\mathbf{P}_{0,1} + \Theta_2^2\mathbf{P}_{0,0} = 0$$

which is the desired result if we identify $\Omega_k^m = -\left(\Theta_1^2\right)^{-1}\Theta_0^2$ and $\mathbf{Q}_k^m = -\left(\Theta_1^2\right)^{-1}\Theta_2^2$. Herewith the relation (15) is shown.

Note that from a computational point of view the matrices $\mathbf{Q}_k^m$ and $\Omega_k^m$ can be computed using a straightforward iteration procedure.

In the computation of performance measures we shall also need information about the generating function of the boundary states $v(\xi)$ at $\xi = 1$. Again due to lack of space in this paper we refer to the paper by Sleptchenko et al. [10] where the procedure of deriving the function $v(\xi)$ at $\xi = 1$ was presented in detail.

### 4.4 States with empty queue ($m = n = 0$).

The equilibrium equations for $\left|\overline{s}^h\right| + \left|\overline{s}^l\right| < k$ do not have an inhomogeneous term:

$$\begin{aligned}
\mathbf{D}_0\mathbf{P}_{0,0} &= \mathbf{E}'_0\mathbf{P}'_{-1} + \mathbf{B}_0\left[\mathbf{Z}^{-1}(\xi)\,\mathbf{C}(\xi)\right]_{\xi=0} + \mathbf{G}_0\mathbf{P}_{0,1}, \quad r = 0 \\
\mathbf{D}'_{-r}\mathbf{P}'_{-r} &= \mathbf{E}'_{-r}\mathbf{P}'_{-r-1} + \mathbf{G}'_{-r}\mathbf{P}_{-r}, \quad 0<r<k \\
\mathbf{D}'_{-k}\mathbf{P}_{-k} &= \mathbf{G}'_{-k}\mathbf{P}'_{-k+1}, \quad r = k
\end{aligned}$$

$$\mathbf{D}'_{-r}\mathbf{P}'_{-r} = \mathbf{E}'_{-r}\mathbf{P}'_{-r-1} + \mathbf{G}'_{-r}\mathbf{P}'_{-r}$$

However, now the matrices $\mathbf{D}'_{-r}$,$\mathbf{E}'_{-r}$ and $\mathbf{G}'_{-r}$ depend on the value of $r$. We can find $\mathbf{P}'_{-r}$ for $0 < r < k$ using a recurrent expression. The complete solution can be represented as:

$$\mathbf{P}'_{-r} = \mathbf{Q}_{-r-1}\mathbf{P}'_{-r-1} = \mathbf{Q}_{-r-1}\mathbf{Q}_{-r-2}\cdots\mathbf{Q}_{-k}\mathbf{P}'_{-k}$$

where $\mathbf{Q}_{-r}$ follows recursively from

$$\mathbf{Q}_{-r-1} = \left(\mathbf{D}_{-r} - \mathbf{G}_{-r}\mathbf{Q}_{-r}\right)^{-1}\mathbf{E}_{-r}, \quad r = 1, \ldots, k$$

In this recurrent relation, the matrix $\mathbf{Q}_0$ was found in the previous section. The free constant $\mathbf{P}'_{-k}$ is determined by

$$\sum_{\overline{w}^h,\overline{s}^h,\overline{w}^l,\overline{s}^l} P_{n,m}\left(\overline{w}^h, \overline{s}^h, \overline{w}^l, \overline{s}^l\right) = 1$$

Hence, we found the probability states for $\hat{t} = 0 \ldots k+1$ and the values of the first derivative of the function $v(1)$ that we need to calculate the performance measures for the queueing system.

## 5 Performance measures.

In this section we present only the performance criteria for the high and low priority items in the queue, since the other performance criteria can be derived following the lines presented by Van Harten and Sleptchenko [9] and by Sleptchenko et al. [10].

The mean number of the low priority item $i$ in the queue can be found as sum of all probability states with low priority items in the queue (i.e. zones I and II) multiplied by the number of low priority items $i$ in the queue:

$$E\left[q_i^*\right] = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} \sum_{\substack{\overline{w}^h, \overline{s}^h \overline{w}^l, \overline{s}^l \\ \left|\overline{w}^h\right| = n, \left|\overline{w}^l\right| = m}} w_i^* P_{n,m}\left(\overline{w}^h, \overline{s}^h, \overline{w}^l, \overline{s}^l\right).$$

In this expression various terms can be simplified via the function $v(\xi)$ and via the matrix $\mathbf{Z}(\xi)$ using Taylor expansion and we have

$$\begin{aligned} E\left[q_i^h\right] &= a_i^h \left\langle \overline{\mathbf{1}}, \left(\mathbf{Z}(1) - \mathbf{I}\right)^{-1} \mathbf{Z}(1)\, v(1) \right\rangle \\ E\left[q_i^l\right] &= -a_i^l \left\langle \overline{\mathbf{1}}, \left(\mathbf{Z}(1) - \mathbf{I}\right)^{-1} \mathbf{Z}'(1) \left(\mathbf{Z}(1) - \mathbf{I}\right)^{-1} v(1) \right\rangle \\ &\quad + a_i^l \left\langle \overline{\mathbf{1}}, \left(\mathbf{Z}(1) - \mathbf{I}\right)^{-1} \mathbf{Z}(1)\, v'(1) \right\rangle + a_i^l \left\langle \overline{\mathbf{1}}, v'(1) \right\rangle \end{aligned}$$

where vector $\overline{\mathbf{1}}$ has all elements equal to 1 and dimension equal to $d\left(N^h + N^l, k\right)$.

## 6 Conclusions and generalizations.

In this paper we derived a method to analyze multi-class $M/M/k$ priority queues with non-preemptive priority and two priority groups (high and low). Each group of priority can contain several classes of items with different arrival and service rates. The proposed method is based on the solution of the stationary state equations. It uses an iteration algorithm.

Also it can be used for cases where items have hyperexponential $(H_r)$ service times. We can deal with these cases by representing each class as $r$ classes with exponentially distributed service time and using the performance estimators for the total number of items in the system among these $r$ classes.

# References


[1] Buzen, J.P., A.B.Bondi. The response times of priority classes under preemptive resume in $M/M/m$ queues. Operations Research. 31, No.3, 456-465, 1983.

[2] Gail, H.R., S.L.Hantler, B.A.Taylor. Analysis of a non-preemptive priority multiserver queue. Advances in Applied Probability. 20, No.4, 852-879, 1988

[3] Gail, H.R., S.L.Hantler, B.A.Taylor. On preemptive Markovian queue with multiple servers and two priority classes. Mathematics Operations Research. 17, No.2, 365-391, 1992

[4] Kanadia, A.S., M.F.Kazmi and A.C.Mitchell. Analysis of a finite capacity non preemptive priority queue. Computers and Operations Research, 11/3, 337-343, 1984.

[5] Kao, E.P.C. and K.S.Narayanan. Computing steady state probabilities of a nonpreemptive priority queue. ORSA Journal on Computing, 2, 211-218, 1990.

[6] Kao, E.P.C., and S.D. Wilson. Analysis of nonpreemptive priority queues with multiple servers and two priority classes. European Journal of Operational Research. 118, 181-193, 1999.

[7] Kella, O., and U. Yechiali. Waiting times in the non-preemptive priority $M/M/c$ queue. Communications in Statistics - Stochastic Models, 1, 257-262, 1985.

[8] Mitrani, I., P.J.B. King. Multiprocessor systems with preemptive priorities. Performance Evaluation 1, 118-125, 1981.

[9] Harten, A. van, A.Sleptchenko. On multi-class multi-server queueing and spare parts management. Queueing systems 43(4), 307-328, 2003.

[10] Sleptchenko, A., A. van Harten, M.C. van der Heijden. On multi-class multi-server queue with preemptive priorities. BETA report (WP-77), University of Twente. Submitted for publication.

[11] Sleptchenko, A., M.C. van der Heijden, A. van Harten. An exact analysis of the multi-class $M/M/k$ priority queue with outsourcing. Submitted for publication.

[12] Wagner, D. Analysis of mean values of a multi-server model with non-preemptive priorities and non-renewal input. Communications in Statistics - Stochastic Models. 13(1), 67-84, 1997.

[13] Wagner, D. Waiting time of a finite-capacity multi-server model with non-preemptive priorities. European Journal of Operational Research. 102, 227-241, 1997.

[14] Wagner, D. A finite-capacity multi-server multi-queueing model with non-renewal input. Annals of Operational Research. 79, 63-82, 1998.