# On the equivalence between multiclass PS-type scheduling policies

Konstantin Avrachenkov
INRIA, Sophia Antipolis

Tejas Bodas
LAAS-CNRS, Toulouse



## ABSTRACT

Consider a single server queue serving a multiclass population. Some popular scheduling policies for such a system (and of interest in this paper) are the discriminatory processor sharing (DPS), discriminatory random order service (DROS), generalized processor sharing (GPS) and weighted fair queueing (WFQ). The aim of this paper is to show a certain equivalence between these scheduling policies for the special case when the multiclass population have identical and exponential service requirements. In fact, we show the equivalence between two broader classes of policies that generalize the above mentioned four policies. We specifically show that the sojourn time distribution for a customer of a particular class in a system with the DPS (GPS) scheduling policy is a constant multiple of the waiting time distribution of a customer of the same class in a system with the DROS (respectively WFQ) policy.

## 1. INTRODUCTION

Consider a single server queue with multiclass customers where the customers have independent and identical service requirements independent of their class. Suppose that there are $N$ customer classes and assume that the service requirement of a customer is exponentially distributed with rate $\mu$. Let $\lambda_i$ denote the arrival rate for a Class $i$ customer, $i = 1 \ldots N$. Let $\sum_{i=1}^{N} \lambda_i = \Lambda$, $\rho_i = \frac{\lambda_i}{\mu}$ and $\rho = \sum_{i=1}^{N} \rho_i$. Further, let $p_i$ denote a weight parameter associated with a Class $i$ customer. Additionally, for the purpose of stability, assume throughout that $\Lambda < \mu$. Some examples of scheduling policies used in such multiclass queues are DPS, DROS, GPS and WFQ. In this paper, we will show that the sojourn time distribution of any Class $i$ customer with DPS (GPS) scheduling policy is equal to the waiting time distribution of a Class $i$ customer in a system with DROS (respectively WFQ) policy under the assumption of i.i.d exponential service requirements. A quick overview of these four scheduling policies is given below.

Scheduling policies such as GPS and DPS are variants of the processor sharing policy and can serve multiple customers from the system simultaneously. In case of GPS, a separate queue is maintained for each customer class and the total service capacity of the server is shared among customers of the different classes in proportion to the weights $p_i$. The GPS scheduling policy is often considered as a generalization of the head-of-line processor sharing policy (HOLPS) as described in [6, 14, 15]. (Refer [16, 17] for details about HOLPS). As a generalization of HOLPS, GPS maintains a FIFO scheduling policy within each queue for a class and only the head-of-line customers of different classes are allowed to share the processor. The share of the server for a head-of-line Class $i$ customer is proportional to the weight $p_i$ and is independent of the number of other customers in the queue. The service rate received by the customer is precisely given by $\frac{p_i}{\sum_{j=1}^{N} p_j \phi_j}$ where $\phi_j = 1$ if the queue has at least one class $j$ customer and $\phi_j = 0$ otherwise. Refer Parekh and Gallager [11], Zhang et al. [12] for an early analysis of the model.

In case of DPS, the total service capacity is shared among all the customers present in the system and not just among the head-of-line customers of different classes. The share of the server for a customer of a class is not only in proportion to the class weight, but also depends on the number of multiclass customers present in the queue. In particular, a Class $i$ customer in the system is served at a rate of $\frac{p_i}{\sum_{j=1}^{N} p_j n_j}$ where $n_j$ denotes the number of Class $j$ customers in the system. The DPS system was first introduced by Kleinrock [5] and subsequently analyzed by several authors [13, 8, 7, 9].

The DROS and WFQ scheduling policies are also characterized by an associated weight for each customer class. However these policies are not a variant of the processor sharing policies and hence their respective server can only serve one customer at a time. The DROS and WFQ policies differ in their exact rule for choosing the next customer. In the DROS policy, the probability of choosing a customer for service depends on the weights and the number of customers of the different classes in the queue. A Class $i$ customer is thus chosen with a probability of $\frac{p_i}{\sum_{j=1}^{N} p_j n_j}$ where $n_j$ denotes the number of Class $j$ customers in the system. DROS policy is also know as relative priority policy and was first introduced by Haviv and wan der Wal [7]. For more analysis of this policy we refer to [3, 10]. In the WFQ policy, a separate queue for each class is maintained and the next customer is chosen randomly from among the head-of-line customers of different classes. As in case of the GPS scheduling policy, a FIFO scheduling policy is used within each queue for a class. WFQ can be seen as a packetised version of GPS and the probability of choosing a head-of-line Class $i$ customer for service is given by $\frac{p_i}{\sum_{j=1}^{N} p_j \phi_j}$ where $\phi_j$ is as defined earlier. Refer Demers [14] for the detailed analysis of the WFQ policy.

It is interesting to note that for a Class $i$ customer, the service rate received *in DPS* and the probability of being chosen next for service *in case of DROS* is given by $\frac{p_i}{\sum_{j=1}^{N} p_j n_j}$. Similarly, the service rate received *in GPS* and the probability of being chosen next for service *in case of WFQ* is $\frac{p_i}{\sum_{j=1}^{N} p_j \phi_j}$. This similarity in the scheduling rules motivates us to compare the mean waiting times and sojourn times of the multiclass customers with these scheduling policies. Having assumed identical service requirements, we specifically show that the waiting time distribution of a Class $i$ customer in a system with DROS (WFQ) scheduling policy is $\rho$ times the sojourn time distribution of any Class $i$ customer with DPS (resp. GPS) scheduling policy. This is a generalization of [2], where the equivalence has been established between single-class processor sharing and random order service discipline. The coupling technique which we use also builds upon the technique used in [2].

The rest of the paper is organized as follows. In the next section, we introduce a generalized notion of multiclass processor sharing (mPS) and random order service (mROS) policies. The DPS, GPS, DROS and WFQ policies will turn out to be special cases of mPS and mROS. In Section 3, we show that the mean sojourn time of a Class $i$ customer with mPS scheduling is equivalent to the mean waiting time of a Class $i$ customer with mROS policy. As a special case, this proves the mentioned equivalences among the four multiclass scheduling policies.

## 2. GENERALIZED MULTICLASS SCHEDULING POLICIES

In this section, we will describe two multiclass scheduling policies that are a generalization of policies such as DPS, DROS, GPS and WFQ. The two policies are based on the processor sharing and random order service mechanism and will be labeled as mPS and mROS respectively.

The mPS scheduling policy is a multiclass processor sharing policy and can serve multiple customers simultaneously. A separate queue for each customer class is maintained and a FIFO sequencing policy is used within each queue of a class. The mPS scheduling policy is parameterized by a vector $\bar{\alpha}$ of length $N$ that characterizes the maximum number of customers of each class that can be served simultaneously with other customers. We shall henceforth use the notation mPS($\bar{\alpha}$) where $\bar{\alpha} = [\alpha_1, \ldots, \alpha_N]$. Here $\bar{\alpha}$ denotes the set of customers that can be served simultaneously. Recall the definition that $n_i$ denotes the instantaneous number of Class $i$ customers in the queue, where $i = 1, \ldots, N$. Let $\beta_i(\bar{n})$ denote the number of Class $i$ customers under service when the configuration of total multiclass customers is $\bar{n}$. Then, clearly $\beta_i(\bar{n}) = \min\,(n_i, \alpha_i)$. To lighten notation, we shall drop the dependence on $\bar{n}$ and use only $\beta_i$ when the context is clear. In other words, if $n_i \le \alpha_i$, then all the Class $i$ customers in the queue are being served simultaneously for $i = 1, \ldots, N$. However if $n_i > \alpha_i$, then only the first $\alpha_i$ customers of Class $i$ in its queue are served simultaneously. Recall that due to the FIFO sequencing policy within each class, only the first $\beta_i$ customers in the queue are always served. For an mPS($\bar{\alpha}$) scheduling policy with a configuration of $\bar{n} = [n_1, \ldots, n_N]$ multiclass customers in the system, the service rate received by a particular Class $i$ customer in service is given by $\frac{p_i}{\sum_{j=1}^{N} p_j \beta_j}$. When $\alpha_i = \infty$, for $i = 1$ to $N$, the corresponding scheduling policy will be denoted by mPS($\bar{\infty}$). In this case, $\beta_i = \min\,(n_i, \infty) = n_i$ and therefore mPS($\bar{\infty}$) corresponds to the DPS scheduling policy. Similarly if $\bar{e} = [1, \ldots, 1]$, then mPS($\bar{e}$) corresponds to the GPS scheduling policy where only the head-of-line customers of each class can be served.

In a similar manner, we can define the mROS($\bar{\alpha}$) scheduling policy where $\bar{\alpha} = [\alpha_1, \ldots, \alpha_N]$ and $\bar{\alpha}$ denotes the set of customers from which the subsequent customer is chosen for service. As in case of the mPS policy, note that a separate FIFO queue for each customer class is also maintained for the mROS system. At any given time, the first $\beta_i = \min\,(n_i, \alpha_i)$ customers are candidates for being chosen for service while the remaining $n_i - \beta_i$ customers have to wait for their turn. For an mROS($\bar{\alpha}$) scheduling policy with a configuration of $\bar{n} = [n_1, \ldots, n_N]$ waiting customers, a Class $i$ customer within the first $\beta_i$ customers in its queue will be chosen next for service with probability $\frac{p_i}{\sum_{j=1}^{N} p_j \beta_j}$. As in case of the mPS scheduling, mROS($\bar{\infty}$) corresponds to the DROS policy whereas mROS($\bar{e}$) corresponds to the WFQ policy.

**Remark** 1. *A policy closely related to the mPS discipline is the limited processor sharing (LPS) policy. LPS is a single class processor sharing policy parametrized by an integer $c$ where $c$ denotes the maximum number of customers that can be served simultaneously. Here $c = \infty$ corresponds to the processor sharing policy while $c = 1$ corresponds to FCFS policy. LPS can also be viewed as a special case of the mPS policy when there is a single service class for the arriving customers. See [1, 18] more more details about the LPS-c policy.*

Having introduced the generalized multiclass scheduling policies, we shall now establish a relation between the sojourn time of a Class $i$ customer in mPS system with the waiting time of a Class $i$ customer in mROS system.

## 3. COMPARING THE SOJOURN AND WAITING TIME DISTRIBUTIONS IN MPS AND MROS

The analysis in this section is inspired from that in [2] where a similar result is established for the case of a single class of population. Let $\bar{n}$ now denote a vector corresponding to the number of customers of each class present in the queueing system at an arrival instant. We have $\bar{n} = (n_1, \ldots, n_N)$ where $n_i$ denotes the number of Class $i$ customers at the arrival instant. Suppose $\sum_{i=1}^{N} n_i = n$. Let random variable $\boldsymbol{S_i}(\bar{\alpha}, \bar{n})$ denote the conditional sojourn time experienced by an arriving Class $i$ customer in an $mPS(\bar{\alpha})$ system when it sees a configuration of $\bar{n}$ customer on arrival. The corresponding unconditional random variable will be denoted by $\boldsymbol{S_i}(\bar{\alpha})$. We shall occasionally use the notation $mPS(\bar{\alpha}, \bar{n})$ to denote the mPS($\bar{\alpha}$) system with $\bar{n}$ customers. Along similar lines, let the random variable $\boldsymbol{W_i}(\bar{\alpha}, \bar{n})$ denote waiting time (time until chosen for service) experienced by an arriving Class $i$ customer in the mROS system conditioned on the fact that it sees a configuration $\bar{n}$ of waiting customers. This system will be often denoted as $mROS(\bar{\alpha}, \bar{n})$ and the unconditional random variable will be denoted by $\boldsymbol{W_i}(\bar{\alpha})$. Let $\mathbb{P}$ and $\mathbb{P}'$ denote the probability distribution of the random variables $\boldsymbol{S_i}(\bar{\alpha}, \bar{n})$ and $\boldsymbol{W_i}(\bar{\alpha}, \bar{n})$ respectively. (The dependence of these distributions on $\bar{n}$

have been supressed for notational convenience.) We now state the main result of this paper.

**Theorem** 1. *$\rho P(\boldsymbol{S_i}(\bar{\alpha}) > t) = P(\boldsymbol{W_i}(\bar{\alpha}) > t)$ for $i = 1, \ldots, N$.*

PROOF. As in [2], our aim is to first provide a coupling $\left(\hat{\boldsymbol{S}}_{\boldsymbol{i}}(\bar{\alpha},\bar{n}), \hat{\boldsymbol{W}}_{\boldsymbol{i}}(\bar{\alpha},\bar{n})\right)$ with the corresponding law denoted by $\hat{\mathbb{P}}$ such that

- $\hat{\boldsymbol{S}}_{\boldsymbol{i}}(\bar{\alpha},\bar{n}) \stackrel{D}{=} \boldsymbol{S_i}(\bar{\alpha},\bar{n})$ and $\hat{\boldsymbol{W}}_{\boldsymbol{i}}(\bar{\alpha},\bar{n}) \stackrel{D}{=} \boldsymbol{W_i}(\bar{\alpha},\bar{n})$
- $\hat{\mathbb{P}}\left(\hat{\boldsymbol{S}}_{\boldsymbol{i}}(\bar{\alpha},\bar{n}) = \hat{\boldsymbol{W}}_{\boldsymbol{i}}(\bar{\alpha},\bar{n})\right) = 1$

The second requirement will help us show that the two distributions $\mathbb{P}$ and $\mathbb{P}'$ are equal. This follows from the coupling inequality

$$\left\|\mathbb{P} - \mathbb{P}'\right\| \leq 2\hat{\mathbb{P}}\left(\hat{\boldsymbol{S}}_{\boldsymbol{i}}(\bar{\alpha},\bar{n}) \neq \hat{\boldsymbol{W}}_{\boldsymbol{i}}(\bar{\alpha},\bar{n})\right). \tag{1}$$

Such a coupling is precisely obtained as follows.

Consider two tagged Class $i$ customers $X$ and $Y$ that arrive to a $mPS(\bar{\alpha},\bar{n})$ and a $mROS(\bar{\alpha},\bar{n})$ system respectively. This means that at the arrival instant of customer X in the $mPS(\bar{\alpha},\bar{n})$ system, there are $n_i$ Class $i$ customers already present in the system. Similarly, at the arrival instant of customer Y in $mROS(\bar{\alpha},\bar{n})$, there are $n_i$ customers of Class $i$ that are waiting for service in the queue. Recall that $\bar{\beta} = [\beta_1,\ldots,\beta_N]$ where $\beta_i$ in the mPS system denotes the number of Class $i$ customers that are receiving service. In the mROS system, $\beta_i$ denotes those (waiting) Class $i$ customers from which the next customer could be chosen. Note that since $\sum_{i=1}^{N} n_i = n$, with the arrival of customer $X$, the $mPS(\bar{\alpha},\bar{n})$ system has $n+1$ customers. Similarly, with the arrival of customer $Y$, the $mROS(\bar{\alpha},\bar{n})$ system has $n+2$ customers of which one customer is in service and the remaining $n+1$ customers (including customer Y) are waiting for service. We will now specify the rule for forming the required coupling. Since the customers can be distinguished by their class index and also the position in their respective queues, we couple the $n+1$ customers in $mPS(\bar{\alpha},\bar{n})$ with the $n+1$ waiting customers in the $mROS(\bar{\alpha},\bar{n})$ system based on their class and queue position. The coupling must be such that the coupled customers belong to the same class and invariably have the same queue position in their respective queues. It goes without saying that the tagged customers $X$ and $Y$ are also coupled. As in [2], we also couple the subsequent arriving customers and let $D_1, D_2\ldots$ denote i.i.d random variables with an exponential distribution of rate $\mu$. These random variables correspond to service times of a customer in service in $mROS(\bar{\alpha})$. At the service completion epoch, pick a pair of coupled customers randomly from the set of $\bar{\beta}$ customers. The random picking is with a distribution such that a Class $i$ pair from the $\bar{\beta}$ customers is chosen with probability $\frac{p_i}{\sum_{j=1}^{N} p_j\beta_j}$. If the chosen pair belongs to Class $k$, then a class $k$ customer departs from the mPS system while such a customer is taken for service in mROS. This process is repeated till the tagged pair $(X,Y)$ leaves the system. Clearly, this joint probability space is so constructed that the random variables $\hat{\boldsymbol{S}}_{\boldsymbol{i}}(\bar{\alpha},\bar{n}) = \hat{\boldsymbol{W}}_{\boldsymbol{i}}(\bar{\alpha},\bar{n})$ $\hat{\mathbb{P}}$–a.s. From Eq. (1), this implies that

$$\boldsymbol{S_i}(\bar{\alpha},\bar{n}) \stackrel{D}{=} \boldsymbol{W_i}(\bar{\alpha},\bar{n}). \tag{2}$$

Now let random variables $\boldsymbol{N}^{mPS}$ (resp. $\boldsymbol{N}_1^{mROS}$) denote the configuration of the total customers (resp. waiting customers in case of mROS system) as seen by a Class $i$ arrival. The subscript 1 in $\boldsymbol{N}_1^{mROS}$ is used to indicate a busy server. We have the unconditional probabilities given by the following.

$$P(\boldsymbol{S_i}(\bar{\alpha}) > t) \quad = \quad \sum_{\bar{n}} P(\boldsymbol{N}^{mPS} = \bar{n}) P(\boldsymbol{S_i}(\bar{\alpha},\bar{n}) > t). \tag{3}$$

Similarly we have

$$P(\boldsymbol{W_i}(\bar{\alpha}) > t) = \sum_{\bar{n}} P(\boldsymbol{N}_1^{mROS} = \bar{n}) P(\boldsymbol{W_i}(\bar{\alpha},\bar{n}) > t). \tag{4}$$

Now if suppose $P(\boldsymbol{N}_1^{mROS} = \bar{n}) = \rho P(\boldsymbol{N}^{mPS} = \bar{n})$ is true, then from Eq. (2), the statement of the theorem follows and this would complete the proof. In the following lemma, we shall prove that indeed $P(\boldsymbol{N}_1^{mROS} = \bar{n}) = \rho P(\boldsymbol{N}^{mPS} = \bar{n})$. $\square$

**Lemma** 1. *$P(\boldsymbol{N}_1^{mROS} = \bar{n}) = \rho P(\boldsymbol{N}^{mPS} = \bar{n})$ for $\bar{n}$ such that $|\bar{n}| \geq 0$.*

PROOF. We first simplify the notations as follows. Let $\pi(\bar{n}) := P(\boldsymbol{N}^{mPS} = \bar{n})$ and $\hat{\pi}(1,\bar{n}) := P(\boldsymbol{N}_1^{mROS} = \bar{n})$. Let $\hat{\pi}(0,\bar{0})$ denote the probability that the mROS system has no customers and is idle. The statement of the lemma now requires us to prove that $\hat{\pi}(1,\bar{n}) = \rho\pi(\bar{n})$. To prove this result, consider the balance equation for the mPS system where $\pi$ shall denote the stationary invariant distribution for the system. The assumption $\Lambda < \mu$ implies that the underlying Markov process is ergodic and hence the stationary distribution $\pi$ is unique. For $\bar{n}$ such that $|\bar{n}| \geq 0$, the detailed balance equations for the mPS($\bar{\alpha}$) system are

$$\begin{aligned} &(\Lambda + \sum_{i=1}^{M}\left(\frac{\beta_i(\bar{n})p_i}{\sum_{j=1}^{M} p_j\beta_j(\bar{n})}\right)\mu\mathbb{1}_{\{|\bar{n}|>0\}})\pi(\bar{n}) \\ = \quad & \sum_{i=1}^{M}\lambda_i\mathbb{1}_{\{n_i>0\}}\pi(\bar{n}-e_i) \\ + \quad & \sum_{i=1}^{M}\left(\frac{\beta_i(\bar{n}+e_i)p_i}{\sum_{j=1}^{M} p_j\beta_j(\bar{n}+e_i)}\right)\mu\pi(\bar{n}+e_i). \end{aligned}$$

Now since

$$\sum_{i=1}^{M}\left(\frac{\beta_i(\bar{n})p_i}{\sum_{j=1}^{M} p_j\beta_j(\bar{n})}\right) = 1,$$

the balance equations can be written as

$$\begin{aligned} (\Lambda + \mu\mathbb{1}_{\{|\bar{n}|>0\}})\pi(\bar{n}) \quad = \quad & \sum_{i=1}^{M}\lambda_i\mathbb{1}_{\{n_i>0\}}\pi(\bar{n}-e_i) \\ + \quad & \sum_{i=1}^{M}\left(\frac{\beta_i(\bar{n}+e_i)p_i}{\sum_{j=1}^{M} p_j\beta_j(\bar{n}+e_i)}\right)\mu\pi(\bar{n}+e_i). \end{aligned} \tag{5}$$

Similarly, the detailed balance equations for the mROS($\bar{\alpha}$) system are as follows for $\bar{n}$ such that $|\bar{n}| \geq 0$.

$$(\Lambda + \mu \mathbb{1}_{\{|\bar{n}|>0\}})\hat{\pi}(1,\bar{n}) = \sum_{i=1}^{M} \lambda_i \mathbb{1}_{\{n_i>0\}} \hat{\pi}(1, \bar{n} - e_i)$$

$$+ \quad \sum_{i=1}^{M} \left( \frac{\beta_i(\bar{n}+e_i)p_i}{\sum_{j=1}^{M} p_j \beta_j(\bar{n}+e_i)}. \right) \mu \hat{\pi}(1, \bar{n}+e_i) \tag{6}$$

Additionally, the idle system should satisfy

$$\Lambda \hat{\pi}(0, \bar{0}) = \mu \hat{\pi}(1, \bar{0}). \tag{7}$$

where $\hat{\pi}(0, \bar{0}) = 1 - \rho$ is the probability that the system is empty.

Now again, the assumption $\Lambda < \mu$ implies that the underlying Markov process is ergodic and hence the stationary distribution $\hat{\pi}$ is also unique. Therefore to prove the lemma, it is sufficient to check if the detailed balance equations for the mROS system given by Eq. (6) are satisfied when $\hat{\pi}(1, \bar{n}) = \rho\pi(\bar{n})$.

Now from Eq. (6) and assuming that $\hat{\pi}(1, \bar{n}) = \rho\pi(\bar{n})$, we have

$$\begin{aligned}
&(\Lambda + \mu \mathbb{1}_{\{|\bar{n}|>0\}})\hat{\pi}(1,\bar{n}) - \sum_{i=1}^{M} \lambda_i \mathbb{1}_{\{n_i>0\}} \hat{\pi}(1, \bar{n} - e_i) \\
&\quad - \sum_{i=1}^{M} \left( \frac{\beta_i(\bar{n}+e_i)p_i}{\sum_{j=1}^{M} p_j \beta_j(\bar{n}+e_i)} \right) \mu \hat{\pi}(1, \bar{n}+e_i) \\
&= (\Lambda + \mu \mathbb{1}_{\{|\bar{n}|>0\}})\rho\pi(\bar{n}) - \sum_{i=1}^{M} \lambda_i \mathbb{1}_{\{n_i>0\}} \rho\pi(\bar{n} - e_i) \\
&\quad - \sum_{i=1}^{M} \left( \frac{\beta_i(\bar{n}+e_i)p_i}{\sum_{j=1}^{M} p_j \beta_j(\bar{n}+e_i)} \right) \mu \rho\pi(\bar{n}+e_i) = 0.
\end{aligned}$$

The last equality follows from Eq. (5) after dividing throughout by $\rho$. Similarly,

$$\begin{aligned}
\Lambda \hat{\pi}(0, \bar{0}) - \mu \hat{\pi}(1, \bar{0}) &= \Lambda \hat{\pi}(0, \bar{0}) - \mu\rho\pi(\bar{0}) \\
&= \mu \left( \rho \hat{\pi}(0, \bar{0}) - \rho\pi(\bar{0}) \right) \\
&= \mu \left( \rho \hat{\pi}(0, \bar{0}) - \rho(1-\rho) \right) \\
&= 0. 
\end{aligned} \tag{8}$$

Here the third equality is from the fact that $\pi(\bar{0}) = (1-\rho)$ is the probability that the mPS($\bar{\alpha}$) system is empty. Clearly, substituting $\hat{\pi}(1, \bar{n}) = \rho\pi(\bar{n})$, satisfies the balance equations for the mROS system. Since $\hat{\pi}$ is the unique invariant distribution, the statement of the lemma follows. □

We now have the following corollary that establishes the desired equivalence between DPS (GPS) and DROS (resp. WFQ) scheduling policies. Note that the result is true only for the case when all customers have identically distributed service requirements. The equivalence result need not be true in general when the customer classes differ in their service requirements.

**Corollary** 1.

- $\rho P(\boldsymbol{S_i}(\bar{\infty}) > t) = P(\boldsymbol{W_i}(\bar{\infty}) > t)$
  *where $\boldsymbol{S_i}(\bar{\infty})$ denotes the sojourn time of a Class $i$ customer in DPS system and $\boldsymbol{W_i}(\bar{\infty})$ denotes the waiting time of a Class $i$ customer in DROS system.*
- $\rho P(\boldsymbol{S_i}(\bar{e}) > t) = P(\boldsymbol{W_i}(\bar{e}) > t)$
  *where $\boldsymbol{S_i}(\bar{e})$ denotes the sojourn time of a Class $i$ customer in GPS system and $\boldsymbol{W_i}(\bar{e})$ denotes the waiting time of a Class $i$ customer in WFQ system.*


# 4. ACKNOWLEDGEMENTS

The authors would like to thank Dr. Rudesindo Queija from CWI, Amsterdam for the several discussions on this work.


# 5. REFERENCES


[1] B. Avi-Itzhak and S. Halfin, "Expected response times in a non-symmetric time sharing queue with a limited number of service positions," *Proceedings of ITC 12*, 1988.

[2] S.C. Borst, O.J. Boxma, J.A. Morrison and R. Núñez-Queija, "The equivalence between processor sharing and service in random order" *Operations Research Letters*, vol. 31, no. 4, pp. 254–262, July 2003.

[3] M. Haviv and J. van der Wal, "Waiting times in queues with relative priorities," *Operations Research Letters*, vol. 35, no. 5, pp. 591–594, September 2007.

[4] M. Haviv and J. van der Wal, "Mean sojourn times for phase-type discriminatory processor sharing systems," *European Journal of Operational Research*, vol. 189, no. 2, pp. 375–386, September 2008.

[5] L. Kleinrock, Time-shared systems: a theoretical treatment. *Journal of ACM*, 14(2):242–261, April 1967.

[6] S. Aalto, U. Ayesta, S. Borst, V. Misra and R. Núñez-Queija, "Beyond Processor Sharing," *ACM Sigmetrics Performance Evaluation Review*, vol. 34, pp. 36–43, 2007,

[7] M. Haviv and J. van der Wal, Equilibrium strategies for processor sharing and random queues with relative priorities. *Probability in the Engineering and Informational Sciences*, null(4):403–412, October 1997.

[8] M. Haviv and J. van der Wal, Waiting times in queues with relative priorities. *Operation Research Letters*, 35:591–594, 2007.

[9] J. Kim, and B. Kim, " Sojourn time distribution in M/M/1 queue with discriminatory processor sharing," *Performance Evaluation*, vol. 58, pp. 341–365, July 2004,

[10] J. Kim, J. Kim and B. Kim, " Analysis of M/G/1 queue with discriminatory random order service policy," *Performance Evaluation*, vol. 68, pp. 256–270, 2011,

[11] A. K. Parekh and R. G. Gallager, A generalized processor sharing approach to flow control in integrated services networks: The single-node case. *IEEE/ACM Transactions on Networking*, 1(3):344–357, 1993.

[12] Z. Zhang, D. Towsley and J. Kurose, Statistical analysis of generalized processor sharing scheduling discipline. In *Proceedings of SIGCOMM*, pages 68–77, 1994.

[13] G. Fayolle, I. Mitrani and R. Iasnogorodski, "Sharing a processor among many job classes," *Journal of ACM*, vol. 27, no. 3, pp. 519–532, July 1980.

[14] A. Demers, "Analysis and Simulation of a Fair Queueing Algorithm," *Internetworking: Research and Experience*, vol. 1, pp. 3–26, 1990.

[15] A. Kumar, D. Manjunath and J. Kuri, " Communication Networking: An Analytical Approach," *Elsevier*, 2004.

[16] G. Fayolle and R. Iasnogorodski, "Two Coupled Processors: The reduction to a Riemann-Hilbert Problem," *Wahrscheinlichkeitstheorie verw Gebiete*, vol. 47, pp. 325–351, 1979.

[17] J. Morrison, "Head of the line processor sharing for many symmetric queues with finite capacity," *Queueing Systems*, vol. 14, pp. 215–237, 1993.

[18] J. Zhang et al, "Diffusion limits of limited processor sharing queues," *The Annals of Applied Probability*, vol. 21, pp. 745–799, 2011.