Scheduling Policies for a Repair Shop Problem

William Kun Liang, Barış Balcıoğlu, Robert Svaluto
University of Toronto,
Department of Mechanical and Industrial Engineering
5 King's College Rd., Toronto, ON M5S 3G8, CANADA,
liang@mie.utoronto.ca, baris@mie.utoronto.ca, robert.svaluto@utoronto.ca

Abstract

In this paper, we analyze a repair shop serving several fleets of machines that fail from time to time. To reduce downtime costs, a continuous-review spare machine inventory is kept for each fleet. A spare machine, if available on stock, is installed instantaneously in place of a broken machine. When a repaired machine is returned from the repair shop, it is placed in inventory for future use if the fleet has the required number of machines operating. Since the repair shop is shared by different fleets, choosing which type of broken machine to repair is crucial to minimize downtime and holding costs. The optimal policy of this problem is difficult to characterize, and, therefore, is only formulated as a Markov Decision Process to numerically compute the optimal cost and base-stock level for each spare machine inventory. As an alternative, we propose the dynamic Myopic($\mathbf{R}$) policy, which is easy to implement, yielding costs very close to the optimal. Most of the time it outperforms the static first-come-first-served, and preemptive-resume priority policies. Additionally, via our numerical study, we demonstrate that repair shop pooling is better than reserving a repair shop for each fleet.

Keywords and Phrases: Spare part inventory control; multiple finite-population queueing systems; static and dynamic repair scheduling; repair shop pooling

1 Introduction

In manufacturing plants, fleets of different types of machines carry out production which involves several stages. These machines fail from time to time or need to undergo maintenance which may obstruct the flow of semi-finished goods on the shop floor on time, and may decrease targeted production levels (**e.g., Wong, Chan, and Chung, 2012, Chung, Chan, and Chan, 2009**). The role that each fleet plays in production affects its size, its downtime costs, as well as the times to failure characteristics the machines experience. For instance, a small fleet may be used for producing a specific and profitable end product, and if some of the machines are down, the company might suffer from significant profit losses. To decrease the fluctuation in production, the repair and maintenance department should not only be agile in its response to failures, but also have a structure – such as keeping spare machines, or pooling repair resources to handle all types of broken machines –, and a repair scheduling policy to protect the company from downtime related costs.

In this paper, we study a system of multiple fleets of machines, and assume that each machine is subject to failures from time to time. Different fleets have different machine types, and machines in the same fleet are assumed to have identical characteristics (i.e., they are of the same type). Therefore, in queueing theory terminology, each fleet is a finite-calling population. The system aims to have a finite number of machines, which is the fleet size, to be in use at all times. When a machine breaks down, it is sent to a repair shop to be fixed. If there are fewer machines in use than the fleet size, the system incurs a fleet specific downtime cost per unit time for each functional machine that the fleet lacks. To decrease downtime costs, at the expense of incurring holding costs, a continuous-review spare machine inventory is kept for each fleet. **This idea is similar to shortening lead times for delivering products in supply chains, as recommended by Kumar et al., 2006.** When a failure occurs, if there is available stock in the inventory, a spare machine is installed without any delay so that downtime cost is not incurred. If there are no available spares in the inventory, upon each failure, the fleet has one less operating (one more down) machine until the repair shop fixes and sends a repaired machine back to the fleet. During this time, the system incurs downtime cost for each down machine. When a repaired machine is received, if there are no down machines of that type, then the repaired machine is put in the spare machine inventory for future use.

Given this framework, by developing queueing based solutions, we address the following questions: (i) Should there be a separate repair shop for each fleet or should a centralized repair shop (CRS) with a higher capacity serve all the fleets? (ii) Given a shared high capacity CRS, which repair scheduling policy can be practically implemented so that the expected downtime and holding costs under that policy remain close to the optimal cost? In addressing these questions, for all alternatives, we model the repair shop as a single server queueing system where broken machines are deemed to be the customers. Since the goal is decreasing downtime and holding costs, the answers to these questions from our research are important to the repair and maintenance departments of large production facilities or mining sites. Our contribution is designing a dynamic myopic policy, namely the Myopic($\mathbf{R}$) policy, which we show to perform close to the optimal repair scheduling policy.

Our problem is an example of a Markovian queueing system with multiple finite calling populations (see Sztrik, 2001, for a comprehensive bibliography on systems with finite populations, **and Basten et al., 2012 and Zorna, Deckroa, and Lehmkuhlb, 1999 for systems assuming infinite calling populations**). If a separate repair shop is allocated for each fleet, the optimal number of spares to be kept can be easily determined using a birth-and-death model, e.g., Taylor and Jackson (1954). However, analyzing the CRS system presents difficulties. The earlier work in the literature on machine interference (or machine repairperson) problems (MIP) (see Haque and Armstrong, 2007, for a recent literature survey) studies the CRS system without any spare machine inventories. Chandra (1986) employs mean value analysis for the first-come-firs-served (FCFS) repair policy for several fleets of machines in a CRS system with no spare machine inventories. Static priorities are also considered among fleets: Chandra (1986) analyzes the non-preemptive priority policy for multiple fleets, and Miller (1981) studies preemptive and non-preemptive policies for two fleets. For the preemptive-resume priority policy between two fleets, one can also employ the methods of Veran (1984), Jaiswal (1968, p. 71,79), and Bitran and Caldentey (2002). When preemption is not allowed, Iravani, Krishnamurthy, and Chao (2007) show that the optimal repair scheduling policy is a simple static non-preemptive priority policy, and provide the sufficient conditions to prioritize the classes correctly. When preemption is allowed, the optimal policy is partially characterized by Iravani and Kolfal (2005), who show, under certain conditions, a static preemptive-resume priority policy is optimal.

There are fewer studies that consider spare part inventories for each fleet. The most relevant one to our work is by Sahba, Balcıoğlu and Banjevic (2012) who assume identical repair rates for all fleets. As explained in Section 3.1, their recursive formulation for the FCFS policy, which they refer to as the RIF (reserved inventory-FCFS) policy, can be used for our problem so long as the repair rates are identical. For non-identical fleet-specific repair rates, the FCFS policy appears to be computationally difficult except for small size problems, as analyzed by Gross and Ince (1981). For the multi-fleet case repaired under the preemptive-resume policy, the extension of the two-fleet model due to Bitran and Caldentey (2002) to multiple fleets by Sahba, Balcıoğlu and Banjevic (2012) can be employed, as we do in Section 3.2. The optimal repair scheduling policy is also unknown and can only be studied with numerical techniques. We formulate the problem as a Markov Decision Process (MDP) in Section 3.3 and apply the value iteration technique for numerical examples in Section 4.

Another stream of research that is relevant to our work is on production scheduling in a flexible manufacturing plant/inventory setting **(for the application of policies such as shortest processing time policy etc., for which the duration of each job is known, see, e.g. Chan and Chan, 2004, Kumar et al., 2008)**. In this area, single-server make-to-stock queues are used as modeling tools to determine optimal base-stock levels of continuous-review finished goods inventories under a given scheduling policy. The fundamental difference of such studies from ours is that customer arrival rates are not state-dependent, but constant. These policies can be static, such as the FCFS or the preemptive-resume priority policies, or can be dynamic, such as the longest-queue policy which is shown to be more cost effective than the FCFS policy by Zheng and Zipkin (1990) and Zipkin (1995). Under dynamic policies, each time a product has to be scheduled for production, the number of pending production orders for each product type is taken into consideration. This can make a dynamic policy more difficult to implement but at the same time more cost saving than a static policy. Index policies are especially popular in determining dynamic scheduling rules in production/inventory problems. Wein (1992) proposes using the $b\mu/h\mu$ rule: when different products are backordered, produce (give priority to) the one that has the largest value of the index $b_k\mu_k$, where μ_k is the service rate and b_k is the backordering cost rate of product k. In case there are no backorders but some inventory levels are below their base-stock levels, give priority to the one that has the lowest value of the index $h_k\mu_k$, where h_k is the holding cost rate of product k. For other variants of $b\mu$-based index policies, we refer the reader to Ha (1997) and Niño-Mora (2006).

Among the dynamic scheduling policies studied in systems with constant arrival rates, we are motivated from the myopic policies which were first studied by Veatch and Wein (1996) and Peña Perez and Zipkin (1997). Given that there are pending production orders for different types of products (customers), a myopic policy computes the cost rate difference the system would have at the end of a possibly random interval (look-ahead time) for each customer class if the next order to produce were from that class. Multiplying each cost rate difference by the corresponding service rate gives an index for each class of customers with pending production orders. Finally, the myopic policy identifies the type of order with the smallest index as the next order to be produced. In the $M/M/1$ make-to-stock queue, Veatch and Wein (1996) propose service times to be the look-ahead time in their Myopic($\mathbf{S}$) policy. Peña Perez and Zipkin (1997) show that their Myopic($\mathbf{T}$) policy, that uses the sojourn time of a class of customers in the $M/M/1$ queue in the absence of other classes of customers, outperforms the Myopic($\mathbf{S}$) policy. Although the optimal scheduling policy is unknown in this setting, de Véricourt, Karaesmen, and Dallery (2000) provide the conditions under which the Myopic($\mathbf{T}$) policy becomes optimal in the $M/M/1$ queue. Both Myopic($\mathbf{S}$) and Myopic($\mathbf{T}$) policies were initially tested ignoring preemption. In the $M/G/1$

queue, Sanajian, Abouee-Mehrizi, and Balcıoğlu (2010) observe that not allowing preemption can increase costs. As a remedy, they propose employing the preemptive-Myopic($\mathbf{T}$) policy which yielded costs very close to the optimal in their numerical examples.

In this paper, we propose the preemptive-Myopic($\mathbf{R}$) policy in choosing the next machine type to repair. Since preemption is incorporated, each time a new broken machine arrives at the repair shop, or the repair of a machine is completed, the Myopic($\mathbf{R}$) policy computes an index, as Myopic($\mathbf{S}$) and Myopic($\mathbf{T}$) policies do, for each fleet with broken machines. It uses the repair time as the look-ahead time. It differs from the Myopic($\mathbf{S}$) policy because it is preemptive, but more importantly the underlying analysis changes radically due to state-dependent customer arrival rates at the repair shop. The state-dependent arrivals, as explained in Section 3.4, prohibits us from considering sojourn time of the Myopic($\mathbf{T}$) policy as the look-ahead time. Even with this limitation, our extensive numerical study presented in Section 4 shows that the Myopic($\mathbf{R}$) outperforms the FCFS and the static preemptive $b\mu$-policy most of the time.

While the main contribution is designing the efficient Myopic($\mathbf{R}$) policy and compare the relative performances of other policies, we additionally assess the value of repair shop pooling as an alternative to dedicating a separate repair shop for each fleet. This question was also addressed by Sahba and Balcıoğlu (2011) assuming identical repair rates in a CRS system operating under the FCFS policy. Sahba, Balcıoğlu and Banjevic (2012) also assess the benefit of a CRS system assuming that fleets served use the same type of critical component and can share inventories. Both studies show that the CRS system is more cost effective than a system reserving repair resources for each fleet. In our problem, since machine types are different, fleets cannot share inventories and repair rates can be fleet specific. Our examples agree with their results, and attest to the benefit of repair shop pooling. **This finding is in accordance with the view of the importance of information sharing and joint decision making in supply chains as exemplified in Wadhwa, Bibhushan, and Chan (2009), Arora, Chan, and Tiwari (2010), Chan and Prakash. (2012), Gumasta, Chan, and Tiwari (2012).**

The rest of the paper is organized as follows. In Section 2, we define our problem. The static and dynamic scheduling policies along with the formulation of the optimal dynamic repair schedule as an MDP are discussed in Section 3. In Section 4, we present our numerical study, which demonstrates that the dynamic preemptive- Myopic($\mathbf{R}$) policy yields costs close to the optimal solution.

2 Preliminaries

In this paper, we consider r fleets of machines, indexed by i, ($i = 1, 2, \ldots, r$) sharing the resources of a single repair shop. The goal is to have N_i machines (referred to as type i machines) to be functional at all times for each fleet i. However, each machine is subject to failures from time to time. Upon failure, the failed machine is sent to a repair shop to be fixed. For a type i machine, times to failure follow an exponential distribution with rate λ_i. To increase the probability of having all N_i machines functional, for each fleet i, a separate continuous-review spare machine inventory is kept, which is operated according to a base-stock policy with a base-stock level S_i. If there is available stock when a failure occurs, a spare type i machine is installed in the fleet without any delay. Therefore, downtime costs are not incurred. If there are no available spares in the inventory, upon each failure, the fleet has one less operating (one more down) machine until the repair shop can fix and send a repaired machine back to the fleet.

Letting $0 \leq W_i(t) \leq N_i$ and $0 \leq I_i(t) \leq S_i$ denote the number of machines in fleet i and spares in its inventory at time t, respectively, we define $A_i(t) = W_i(t) + I_i(t)$, which gives the number of functional machines (in use or stock) at time t. When $A_i(t) < N_i$, the fleet has $N_i - A_i(t)$ down machines, and the system incurs a downtime penalty cost of b_i per down machine per unit time. Similar to Louit et al. (2011), we assume that the total inventory holding cost to be paid is $h_i \times S_i$ per unit time, since this is the capital cost tied up for keeping S_i additional units of type i machines. One can consider the warehousing cost for the items kept in inventories as well, but as indicated by Waters (2003, p. 257), the capital cost dominates the warehousing costs, and thus, we ignore the latter.

We consider two alternatives regarding the repair shop, which is modeled as a single server queueing system with exponential repair times. In the first, each fleet has its own repair shop. Following Sahba, Balcıoğlu and Banjevic (2012), we refer to this system as the *base case* (BC) system. In the BC system, the repair shop for fleet i is a single server queueing system where the server has a repair rate of μ_i^{BC}. In the second, there is a centralized repair shop (CRS) – again modeled as a single server – that serves all fleets. The CRS has a higher capacity than individual repair shops of the BC system and fixes type i machines at a repair rate of μ_i. In CRS systems, it is assumed that there is no set-up time/cost when the repair shop switches from one type of machine to another. In both systems, if $A_i(t) < N_i$, a repaired type i machine is installed in its fleet right away, raising $W_i(t)$ by 1; otherwise, it is placed in the respective spares inventory raising $I_i(t)$ by 1.

Given this, if one obtains the steady-state probabilities $p_i(n) = P(A_i = n)$, the optimal objective value C^*_{BC} of the BC system cost can be expressed as follows:

$$C^*_{BC} = \sum_{i=1}^{r} C_i(S_i^*), \tag{1}$$

where

$$C_i(S_i^*) \ = \ \min_S \{C_i(S) = h_i \times S + b_i \sum_{n=0}^{N_i} (N_i - n) p_i(n)\}. \tag{2}$$

Note that in a BC system, fleets and their inventories are independent of each other because they have their own repair shops. This makes $p_i(n)$ also independent of other fleets and their spare part inventories. Given S_i for fleet i, these probabilities can be derived by a simple birth-and-death process (e.g., Gross and Harris, 1998, p. 82-83). Therefore, for each fleet, S_i^* is found by searching over different S_i values in Eq. (2). Finally, Eq. (1) gives the optimal BC system cost.

In the CRS system, $p_i(n)$ not only depends on the scheduling policy considered, but also on the characteristics of other fleets due to the shared repair shop. Letting $\mathbf{S} = (S_1, S_2, \ldots, S_r)$, under a given scheduling policy H, the optimal objective value C^*_H of the

CRS system cost can be expressed as follows:

$$C_H^* = \min_{\mathbf{S}}\{\sum_{i=1}^{r} C_i(\mathbf{S})\}, \tag{3}$$

where

$$C_i(\mathbf{S}) \;=\; h_i \times S_i + b_i \sum_{n=0}^{N_i} (N_i - n) p_i(n). \tag{4}$$

Therefore, using Eq. (4) in Eq. (3), $p_i(n)$ values and C_H^* can be found by searching over different $\mathbf{S}$ vectors.

3 Scheduling Policies

In CRS systems where different types of broken machines (jobs) compete for sharing the same repair resource, scheduling policies are needed to determine which machine to fix next. Obviously, the FCFS policy appears as the most straightforward to implement and fair repair shop scheduling policy. However, certain fleets can be more important for a company if they are used to manufacture products with a higher profit (**see Sundarraj, 2006, that explores how customer classes can be prioritized by considering contract and customer type features**). In order to operate these fleets as much as possible with all its machines functional, priority can be given to these types of machines. Under priority policies (preemptive-resume priority policy, in specific, which we consider in this paper), if the inventory level of a higher priority fleet is below its base-stock level, even when a lower priority fleet has some or all machines down, the next machine to fix is from the higher priority fleet.

Irrespective of how many repair jobs there are in the repair shop, the scheduling rule does not change under the FCFS or preemptive-resume priority policies. Dynamic scheduling policies, on the other hand, consider how many jobs are present from each fleet at a given time. The optimal policy, which minimizes the total average cost per unit time, for this problem is unknown, yet, it can be formulated as an MDP as presented in Section 3.3. In this paper, we propose the dynamic Myopic($\mathbf{R}$) policy in Section 3.4, which we demonstrate via numerical examples, to outperform the FCFS and preemptive-resume priority policies most of the time, and to result in costs close to the optimal.

3.1 The First-Come-First-Served Policy

Under the FCFS policy, jobs are repaired based on the order of their arrival at the repair shop. Unlike the FCFS multi-class $M/M/1$ queue, where job arrival rate for each class of customers is constant, as also pointed out by Kelly (1975), the exact analysis of the FCFS multi-class $M_n/M/1$ queue with state-dependent arrival rates is difficult. In the $M_n/M/1$ queue with class-specific service rates, as in our case, for state description, not only do we have to know how many jobs there are from each class, but also their ordering in the queue. This makes an analysis based on a continuous-time Markov chain (CTMC) intractable when the problem size increases. Exact solution for two fleets – ignoring spares inventories – is provided by Gross and Ince (1981) for small size fleets. They also propose approximations for larger fleet sizes which turned out to be inaccurate for our examples, thus, are not considered in this study. This leaves us to consider identical repair rates for all fleets for which the exact solution is obtained by Sahba, Balcıoğlu, and Banjevic (2012). In this model, which they refer to as the RIF policy, they show that the ordering of jobs is unimportant and the vector of jobs present in the repair shop is sufficient to recursively obtain $p_i(n)$. Therefore, using their RIF model, the performance of the FCFS policy is assessed in Section 4 only for identical repair rates.

3.2 The Preemptive-Resume Priority Policy

We consider a preemptive-resume priority rule according to which a high-priority job does not have to wait for the completion of the repair of a lower priority job that it sees under repair upon its arrival at the repair shop. If there are no high-priority jobs left in the system, the preempted job resumes its repair from the point of interruption. Within each priority class, the order of service follows the FCFS rule.

Under the preemptive-resume priority policy, $p_i(n)$ can be obtained using models from the literature. Bitran and Caldentey (2002) propose a matrix-geometric form solution to obtain $p_i(n)$ for a queueing system with two classes of customers whose arrival rates are state dependent. For more than two fleets, Sahba, Balcıoğlu, and Banjevic (2012) extend the method of Bitran and Caldentey (2002). In numerical examples in Section 4, therefore, we exploit these two papers for the preemptive-resume priority policy.

3.3 Markov Decision Process Formulation

We formulate the optimal dynamic repair schedule as an MDP similar to Iravani and Kolfal (2005) who analyze the problem without spares inventories, when preemption is allowed. The MDP is formulated as follows:

- *State Space:* State space $\mathcal{S}$ consists of r-dimensional row vectors $\mathbf{n} = (n_1, \ldots, n_r)$ where $0 \leq n_i \leq N_i + S_i$ represents the number of functional type i machines (in use or stock) in the system.
- *Decision Epochs:* Decision epochs are failure instants of a machine from any fleet (i.e., an arrival of a job at the repair shop) and repair completions.
- *Action Set:* For any $\mathbf{n} \in \mathcal{S}$, the set of allowable actions $\mathbf{A_n}$ consists of *Idling* and *Repairing* type i machine if $n_i < N_i + S_i$, $i = 1, \ldots, r$. Therefore, the action set is $\mathcal{A} = \cup_{\mathbf{n} \in \mathcal{S}} \mathbf{A_n}$.

We define $\mathbf{I}^i_{\{a\}}$ as follows

$$\mathbf{I}^i_{\{a\}} = \begin{cases} \mathbf{e}^i & \text{if } a \text{ is true,} \\ \mathbf{0} & \text{otherwise,} \end{cases}$$

where $\mathbf{0}$ is an r-dimensional zero vector, and $\mathbf{e}^i$ is an r-dimensional row vector with 1 on its ith entry and 0 elsewhere. Furthermore, let $\mathbf{J_n}$ denote the set of machine types of which $n_i < N_i + S_i$, i.e., the machine types from which there are broken machines in the repair shop. With $(x)^+ := \max\{0, x\}$, $(N_i - n_i)^+$ gives the number of down machines. Then, the optimality equation for the MDP problem is expressed as follows:

$$\begin{aligned}\frac{g}{\Lambda} + V(\mathbf{n}) &= \frac{1}{\Lambda}\{\sum_{i=1}^{r}\left(h_i S_i + b_i(N_i - n_i)^+\right) + \sum_{i=1}^{r}\min\{N_i, n_i\}\lambda_i V(\mathbf{n} - \mathbf{I}^i_{\{n_i>0\}}) \\ &\quad + \sum_{i=1}^{r}(N_i - n_i)^+\lambda_i V(\mathbf{n}) + f(\mathbf{n})\},\end{aligned} \tag{5}$$

where $V(\mathbf{n})$ **is the value function for the vector** $\mathbf{n}$, $\Lambda = \sum_{i=1}^{r}(N_i\lambda_i + \mu_i)$, and

$$f(\mathbf{n}) = \min\begin{cases}\sum_{i=1}^{r}\mu_i V(\mathbf{n}) & \text{Idling,} \\ \min_{j\in\mathbf{J_n}}\left\{\mu_j V\left(\mathbf{n}+\mathbf{e}^i\right) + \sum_{i=1,i\neq j}^{r}\mu_i V(\mathbf{n})\right\} & \text{Repair,}\end{cases}$$

where g is the total average cost per unit time.

For the MIP in which $S_i = 0$, $i = 1, \ldots, r$, that is, no spares are kept for any fleet, Iravani and Kolfal (2005) define the following two conditions, **C1** or **C2**, for type j and k machines:

Condition **C1**: $b_j\mu_j \geq b_k\mu_k$,

Condition **C2**: $b_j\mu_j < b_k\mu_k$, $\lambda_j < \lambda_k$, and $b_j\mu_j \geq b_k\mu_k\left(1 - (\lambda_k - \lambda_j)/\Lambda\right)$.

In their Theorem 1, they prove that if either Condition **C1** or **C2** holds at a state $\mathbf{n}$ for type j and $k(\neq j)$ machines $(j, k \in \mathbf{J_n})$, then it is optimal to repair type j machine instead of type k machine if

$$\frac{b_j\mu_j}{\lambda_j} \geq \frac{b_k\mu_k}{\lambda_k}. \tag{6}$$

If Conditions **C1** and **C2** are not satisfied in the MIP (with $S_i = 0$, $i = 1, \ldots, r$) the optimal policy is not known. This is also true in our problem where $S_i \geq 0$. However, the optimal average total cost of the model presented in Eq. (5) can be found numerically as we do in Section 4 while computing the optimal system costs.

3.4 The Myopic(R) Policy

The idea of a dynamic myopic policy is to look at the end of a possibly random interval (look-ahead time) in the future and compute the cost rate difference the system would have at that instant if we were to decide repairing a type i machine now, instead of not repairing it, for $i = 1, ..., r$. Since the holding cost rate is constant according to Eq. (4), the machine type that has the largest cost reduction (if the number of down machines is decreasing) is scheduled for repairing next. We consider the look-ahead time to be the repair time, and refer to this policy as the Myopic($\mathbf{R}$) policy. In the production-inventory setting, where customer arrival rates are assumed to be constant, analogous to the Myopic($\mathbf{R}$) policy is the Myopic($\mathbf{S}$) policy for which the look-ahead time is a service time. The Myopic($\mathbf{S}$) policy in a non-preemptive fashion was first studied by Veatch and Wein (1996). Longer look-ahead times such as the sojourn time of a job in a single-product single server queue are also considered, as in the Myopic($\mathbf{T}$) policy by Peña Perez and Zipkin (1997). In the repair shop setting, considering sojourn time as the look-ahead time is technically quite difficult as will be explained later. Additionally, as the numerical examples in Section 4 demonstrate, the Myopic($\mathbf{R}$) policy already yields costs close to the optimal. Thus, we do not consider longer look-ahead times than the repair time.

Since our aim is to design the Myopic($\mathbf{R}$) policy to perform as closely as possible to the optimal policy, the decision epochs are the same as those considered in Section 3.3: failure instants of a machine from any fleet (i.e., an arrival of a job at the repair shop) and repair completions. This implies that the Myopic($\mathbf{R}$) policy may choose to preempt an ongoing repair of a job in favor of another one. The preempted job resumes its repair from the moment of interruption later on when the the Myopic($\mathbf{R}$) policy determines that its type of machine should be repaired next. The better performance of preemptive-resume myopic policies over non-preemptive ones was first discussed by Sanajian, Abouee-Mehrizi, and Balcıoğlu (2010) around the Myopic($\mathbf{T}$) policy in a production/inventory system. This also supports our idea of allowing preemption for the Myopic($\mathbf{R}$) policy in our problem.

Now that the decision epochs are determined, the next step is to compute the cost rate difference at the end of the look-ahead time, which enables us to make the repair scheduling decision. Denoting the inventory position before making the decision by x_i, when $x_i > 0$, one computes the difference between expected cost-rates for type i machines as

$$\Delta c_i^R(x_i) = c_i^R(x_i+1) - c_i^R(x_i),$$

where $c_i^R(x_i) = h_i S_i + b_i \sum_{m=x_i+1}^{x_i+N_i} (n-x_i) p_i^R(m)$. Here, $p_i^R(m)$ is the probability of having m type i machine failures over the look-ahead time. Observe that whether we decide to repair or not, the number of type i machines that can fail over the look-ahead time cannot exceed $x_i + N_i$, i.e., $0 \leq m \leq x_i + N_i$. Then, we have

$$\begin{aligned}\Delta c_i^R(x_i) &= h_i S_i + b_i \sum_{m=x_i+2}^{x_i+N_i} (m - x_i - 1) p_i^R(m) - h_i S_i - b_i \sum_{m=x_i+1}^{x_i+N_i} (m-x_i) p_i^R(m), \\ &= -b_i \sum_{m=x_i+1}^{N_i+x_i} p_i^R(m). \end{aligned} \tag{7}$$

If we have to make the repair scheduling decision at time t and $x_i = 0$, either N_i or fewer machines might be working, i.e., $W_i(t) \leq N_i$. If we choose not to repair a type i machine, the number of down type i machines will increase by m at the end of the look-ahead time, if $0 < m \leq W_i(t)$ more machines fail. If we choose to repair, the number of down machines will increase only by $m-1$. Therefore, we find $\Delta c_i^R(0) = -b_i$. This agrees with $\Delta c_i^S(0) = \Delta c_i^T(0) = -b_i$ of the Myopic($\mathbf{S}$) and Myopic($\mathbf{T}$) policies, respectively, in production/inventory systems.

According to the Myopic($\mathbf{S}$)/Myopic($\mathbf{T}$) policies, not only the amount of $\Delta c_i^S(x_i)$ /$\Delta c_i^T(x_i)$ but also how quickly the server processes a job is an important consideration. Therefore, for each customer class i, the index $\mu_i \Delta c_i^S(x_i)/\mu_i \Delta c_i^T(x_i)$ is computed and the product with the lowest index is scheduled next for production.

In production/inventory systems, the $b\mu$ rule (the static preemptive-resume policy prioritizing the class with the highest $b\mu$ index) is optimal when the inventory levels are 0. Since $\mu_i \Delta c_i^S(0) = \mu_i \Delta c_i^T(0) = -\mu_i b_i$, the Myopic($\mathbf{S}$) and Myopic($\mathbf{T}$) policies turn out to make the optimal scheduling decision when customers are backordered. In a similar vein, we propose using $\mu_i \Delta c_i^R(x_i)/\lambda_i$ as the index so that the machine type with the lowest $\mu_i \Delta c_i^R(x_i)/\lambda_i$ value should be repaired next. Note that when $x_i = 0$, this index equals $-\mu_i b_i/\lambda_i$ and if Conditions

C1 or **C2** determined by Iravani and Kolfal (2005) holds, the Myopic($\mathbf{R}$) policy also makes the optimal repair decision as outlined in Eq. (6).

We now proceed with obtaining $p_i^R(m)$ to compute Eq. (7), which is provided in the following Proposition.

Proposition 1 *The probability of having m failures in fleet i during the look-ahead time depends on x_i, and*

- *if $0 \le m \le x_i$, then*

$$p_i^R(m) = \left(\frac{\lambda_i N_i}{\lambda_i N_i + \mu_i}\right)^m \left(1 - \frac{\lambda_i N_i}{\lambda_i N_i + \mu_i}\right), \tag{8}$$

- *if $0 \le x_i \le m$, then*

$$p_i^R(m) = \frac{\frac{N_i!}{(N_i - m + x_i)!} \frac{\mu_i}{\lambda_i} \frac{(N_i \lambda_i)^{x_i}}{(N_i \lambda_i + \mu_i)^{x_i}}}{\prod_{j=0}^{m-x_i} (N_i + \frac{\mu_i}{\lambda_i} - j)}. \tag{9}$$

Proof. If $0 \le m \le x_i$, then we have enough spares to replace failed machines and times between replacements are exponentially distributed with rate $\lambda_i N_i$. Given that the repair time is $R_i = r$, the number of failures follows Poisson distribution with rate $\lambda_i N_i r$. Removing the condition on r gives $p_i^R(m)$ as

$$\begin{aligned} p_i^R(m) &= \int_0^\infty \frac{(\lambda_i N_i r)^m}{m!} e^{-\lambda_i N_i r} \mu_i e^{-\mu_i r} dr = \frac{\mu_i (\lambda_i N_i)^m}{m!} \int_0^\infty r^m e^{-r(\lambda N + \mu)} dr \\ &= \frac{\mu_i (\lambda_i N_i)^m}{m!} \frac{m!}{(\lambda_i N_i + \mu_i)^{m+1}}, \end{aligned}$$

from which Eq. (8) is obtained.

If $0 \le x_i \le m$, given that the repair time is $R_i = r$, the first x_i failures should happen by time $y < r$, which follows an x_i-stage Erlang distribution in which each exponential phase has a rate of $\lambda_i N_i$. After y, in the remaining $r - y$ time units, $m - x_i$ out of N_i machines should fail and others survive, in order to have a total of n machine failures during r. Then, removing the condition on r, we have

$$\begin{aligned} p_i^R(m) &= \binom{N_i}{m-x_i} \int_0^\infty \int_0^r \left(1 - e^{-\lambda_i (r-y)}\right)^{m-x_i} \left(e^{-\lambda_i (r-y)(N_i - m + x_i)}\right) \frac{(N_i \lambda_i)^{x_i} y^{x_i - 1} e^{-\lambda_i N_i y}}{(x_i - 1)!} dy \mu_i e^{-\mu_i r} dr, \\ &= \binom{N_i}{m-x_i} \int_0^\infty \frac{(N_i \lambda_i)^{x_i} y^{x_i-1} e^{-(\lambda_i N_i + \mu_i) y} \mu_i}{(x_i - 1)!} \left(\int_0^\infty \left(1 - e^{-\lambda_i \nu}\right)^{m-x_i} e^{-\lambda_i \nu (N_i - m + x_i + \frac{\mu_i}{\lambda_i})} d\nu\right) dy, \end{aligned} \tag{10}$$

where $\nu = r - y$. Letting $t = e^{-\lambda_i v}$, $dt = -\lambda_i e^{-\lambda_i v} dv$, $dv = -dt/\lambda_i t$,

$$\int_0^\infty \left(1 - e^{-\lambda_i \nu}\right)^{m-x_i} e^{-\lambda_i \nu (N_i - m + x_i + \frac{\mu_i}{\lambda_i})} d\nu = \frac{1}{\lambda_i} \int_0^1 (1-t)^{m-x_i} t^{N_i - m + x_i + \frac{\mu_i}{\lambda_i} - 1} dt.$$

Since $B(\alpha, \beta) = \int_0^1 (1-t)^{\alpha-1} t^{\beta-1} dt = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha+\beta)$,

$$\begin{aligned}
\int_0^\infty \left(1 - e^{-\lambda_i \nu}\right)^{m-x_i} e^{-\lambda_i \nu (N_i - m + x_i + \frac{\mu_i}{\lambda_i})} d\nu &= \frac{1}{\lambda_i} \frac{\Gamma(m - x_i + 1)\Gamma(N_i - m + x_i + \frac{\mu_i}{\lambda_i})}{\Gamma(N_i + 1 + \frac{\mu_i}{\lambda_i})} \\
&= \frac{1}{\lambda_i} \frac{(m - x_i)!}{(N_i + \frac{\mu_i}{\lambda_i})(N_i + \frac{\mu_i}{\lambda_i} - 1) \cdots (N_i - m + x_i + \frac{\mu_i}{\lambda_i})}.
\end{aligned}$$

Substituting this in Eq. (10) gives,

$$\begin{aligned}
p_i^R(m) &= \frac{\mu_i}{\lambda_i} \frac{\binom{N_i}{m-x_i}(m-x_i)!}{\prod_{j=0}^{m-x_i}(N_i + \frac{\mu_i}{\lambda_i} - j)} \frac{(N_i\lambda_i)^{x_i}}{(x_i - 1)!} \int_0^\infty y^{x_i - 1} e^{-(\lambda_i N_i + \mu_i)y} dy \\
&= \frac{\mu_i}{\lambda_i} \frac{\binom{N_i}{m-x_i}(m-x_i)!}{\prod_{j=0}^{m-x_i}(N_i + \frac{\mu_i}{\lambda_i} - j)} \frac{(N_i\lambda_i)^{x_i}}{(x_i - 1)!} \frac{(x_i - 1)!}{(N_i\lambda_i + \mu_i)^{x_i}},
\end{aligned}$$

from which Eq. (9) is obtained. ■

With Proposition 1 and Eq. (7), at decision epochs, we are now able to compute $\mu_i \Delta c_i^R(x_i)/\lambda_i$ for each fleet i. If we were to consider the sojourn time as the look-ahead time in parallel to the Myopic(**T**) policy, Eq. (10) could be evaluated only by some numerical approximation technique. Additionally, the density function of the sojourn time would change for each base-stock level to consider due to state-dependent arrival rates depending on the base-stock level as well. This would make the search on the optimal base-stock levels extremely long, rendering the policy unpractical.

For a given vector S_i's, $i = 1, \dots, r$, we construct an CTMC for the Myopic(**R**) policy with the same state space $\mathcal{S}$ of the MDP defined in Section 3.3. For each state $\mathbf{n} = (n_1, \dots, n_r)$ except for the state in which $n_i = N_i + S_i$ for all i, let $\mu_{\mathbf{n}}$ denote the repair rate for the type of machine identified by the the Myopic(**R**) policy to be fixed. Using this and the state-dependent failure rates, the global balance equations of the CTMC can be obtained, and $p_i(n)$ can be computed. Searching over different **S** vectors we arrive at the optimal objective value given in Eq. (3).

4 Numerical Results

In Section 3, we propose different repair scheduling policies that can be implemented in a CRS system as alternatives to the BC system discussed in Section 2. However, we did not address several important questions. (i) Although the benefit of server capacity pooling is well-known in the literature of production/inventory systems (see Yu, Benjaafar, and Gerchak, 2009 and the references therein), and repair shop pooling with identical repair rates (see Sahba, Balcıoğlu and Banjevic, 2012, Sahba and Balcıoğlu, 2011) what is the benefit of repair shop pooling when repair rates for different machines are different? (ii) How do the FCFS, the static preemptive-resume priority and the Myopic($\mathbf{R}$) policies perform with respect to one another and the optimal policy? The answers to these questions are important for the management of a repair shop. Instead of reserving its repair resources separately for each fleet, possible benefits from repair shop pooling could be important.

Before presenting our extensive numerical study in detail, we summarize our findings. In regards to Question (i), the results in Section 4.2 clearly demonstrate the advantage of pooling the repair shop capacity. All policies studied in Section 3 result in less cost than the BC system. In regards to Question (ii), the results in Section 4.3 demonstrate that the Myopic($\mathbf{R}$) outperforms the FCFS policy in all, and the static preemptive-resume priority policies in most of the cases, and results in costs very close to the optimal costs.

4.1 Basic Experimental Setup

In order to investigate the questions raised at the beginning of Section 4, we have designed a series of numerical experiments involving two fleets. For the preemptive-resume priority policy, the fleet that has a higher $b_i\mu_i/\lambda_i$ index, $i = 1, 2$, is considered the high-priority class. Hence, we refer to it as the preemptive-$b\mu/\lambda$ policy. This way, we assess the performance the preemptive-$b\mu/\lambda$ policy in settings where it is not known to be optimal, namely, when $S_i > 0$ and Conditions **C1** or **C2** may not hold. For fleet i with N_i machines having λ_i failure rate, we assume μ_i^{BC} to be the repair rate in the BC system, and $\mu_i = 2\mu_i^{BC}$ in the

CRS system.

We considered the following parameters:

- Setting $h_1 = 1$, we consider the following holding cost rates for fleet 2: $h_2 \in \{0.9, 0.7, 0.5\}$.
- We consider the following downtime cost rate to holding cost rate ratios: $b_1/h_1 = b_2/h_2 \in \{20, 80\}$.
- The fleet sizes are : $(N_1, N_2) \in \{(10,5), (10,10), (10,15), (50,25), (50,50), (100,50)\}$.
- When $N_1 = 10, 100$ ($N_1 = 50$), we set $\mu_1 = 2$ ($\mu_1 = 1$), and we consider $\mu_1/\mu_2 \in \{2, 1, 2/3\}$
- As an approximate measure of the repair shop utilization, we set $u = \lambda_1 N_1/\mu_1 = \lambda_2 N_2/\mu_2 \in \{0.45, 0.35, 0.25\}$ corresponding to high, medium, and low levels of repair shop utilization.

Thus, due to three different holding cost rates for fleet 2, two different downtime cost rate to holding cost rate ratios, six different pairs of fleet sizes, three different repair rate ratios, and three different utilizations, we have $3 \times 2 \times 6 \times 3 \times 3 = 324$ cases. For each problem, the optimal base-stock levels for the BC system are obtained from Eq. (2) by searching over different S_i values and the optimal cost of the system is computed from Eq. (1). For each alternative policy in the CRS system, optimal values are are obtained by searching over different $\mathbf{S}$ vectors using Eqs. (3) and (4). However, for the FCFS policy, we only consider the cases with equal repair rates, i.e., when $\mu_1/\mu_2 = 1$ since exact methods or simulation take prohibitively long computation times to handle fleet-specific unequal repair rates. This gives a total of 108 problems for the FCFS policy.

To obtain the optimal costs and base-stock levels as reference values, the value-iteration algorithm as described by Tijms (2003, p. 285) to evaluate Eq. (5) was implemented in C++ and run using a 64-bit compiler from Microsoft Visual Studio Ultimate on a Windows-based computer with Intel i7 CPU and 6.0 GB RAM. The value iteration algorithm was terminated once five-digit accuracy was obtained. The computation time varied depending on the server utilization. When the server utilization was low, i.e. $u = 0.25$, the solution algorithm converged significantly faster compared to problems with higher server utilization. With large fleet sizes, $(N_1, N_2) = (100, 50)$, and high server utilization, $u = 0.45$, the algorithm converged approximately in 2 hours for each problem instance, but at low server utilization, $u = 0.25$, it took approximately 30 minutes to converge for each problem. Even with two-fleet problems, the computation time is considerably long. Therefore, we have chosen not to include problems with more than two flees in our numerical study. **In contrast, the Myopic(R) policy takes the decisions by computing the index for each fleet, and involves solving for the transition probability matrix of the underlying CTMC to obtain the required steady-state probabilities as explained in Section 3.4. Thus, the computation time for any instance was less than a minute.**

4.2 Benefits of Repair Shop Pooling

To answer Question (i) in addressing the benefit of repair shop pooling, for each of the 324 examples (108 for the FCFS policy), we denote the optimal costs by C^*_{BC} for the BC system, and C^*_F for the FCFS policy, C^*_P for the preemptive-$b\mu/\lambda$ policy, C^*_R for the Myopic($\mathbf{R}$) policy in the CRS system. Denoting the optimal cost in the CRS system by C^*, we define

$$\Delta^F_{BC} \equiv \frac{C^*_{BC} - C^*_F}{C^*_{BC}},\ \Delta^P_{BC} \equiv \frac{C^*_{BC} - C^*_P}{C^*_{BC}},\ \Delta^R_{BC} \equiv \frac{C^*_{BC} - C^*_R}{C^*_{BC}},\ \Delta^O_{BC} \equiv \frac{C^*_{BC} - C^*}{C^*_{BC}}.$$

These ratios measure the cost decrease due to repair shop pooling in CRS systems operating under the policies introduced in Section 3 with respect to the optimal BC system.

Table 1 summarizes the cost reduction as a result of repair shop pooling. We see remarkable cost savings under each policy in a CRS system. Sahba, Balcıoğlu, and Banjevic (2010) conducted a similar analysis by assuming the same repair rate for all fleets in a repair shop/inventory system. Their RIF and RIP systems correspond to the FCFS and the preemptive-$b\mu/\lambda$ policies we consider in this paper, respectively. Our findings agree with their observations for these two policies. Results in Table 1 suggest that a CRS system with

Table 1: The minimum, average, median and maximum values of cost reduction due to repair shop pooling.

	Min(%)	Average(%)	Median(%)	Max(%)
Δ_{BC}^{F}	13.85	38.96	38.95	43.77
Δ_{BC}^{P}	6.18	40.37	40.21	63.70
Δ_{BC}^{R}	38.07	46.66	45.6	64.35
Δ_{BC}^{O}	38.07	47.01	45.84	64.44

higher capacity operating under a dynamic repair shop policy is more cost effective than allocating separate repair shops for each fleet.

4.3 Relative Performance of Policies

To answer Question (*ii*) in comparing the relative performance of the policies with one another and the optimal cost, we define

$$\Delta_F^O \equiv \frac{C_F^* - C^*}{C_F^*},\ \Delta_F^P \equiv \frac{C_F^* - C_P^*}{C_F^*},\ \Delta_F^R \equiv \frac{C_F^* - C_R^*}{C_F^*},$$

and

$$\Delta_P^O \equiv \frac{C_P^* - C^*}{C_P^*},\ \Delta_P^R \equiv \frac{C_P^* - C_R^*}{C_P^*},\ \Delta_R^O \equiv \frac{C_R^* - C^*}{C_R^*}.$$

The ratios Δ_F^P, Δ_F^R, and Δ_P^R measure the relative performances of the non-optimal policies considered for a CRS system. Table 2 summarizes the cost reduction in pairwise comparison of non-optimal policies (a negative number indicates a cost increase). We conclude that the Myopic($\mathbf{R}$) policy is superior to the FCFS policy in all the examples considered. In general, it also reduces the system cost considerably when compared to the preemptive-$b\mu/\lambda$ policy, however, in some cases the latter is more cost effective than the former. We also do not see that the preemptive-$b\mu/\lambda$ policy is always better than the FCFS policy in reducing the system cost.

Table 2: The minimum, average, median and maximum values of cost reduction in pairwise comparison of non-optimal policies.

	Min(%)	Average(%)	Median(%)	Max(%)
Δ_F^P	-15.85	3.38	1.95	33.95
Δ_F^R	4.39	11.95	10.43	35.1
Δ_P^R	-2.97	9.48	8.19	46.52

The ratios Δ_F^O, Δ_P^O, and Δ_R^O measure the cost decrease of the optimal policy when compared to the three non-optimal policies.

Table 3: The minimum, average, median and maximum values of cost reduction due to the optimal policy.

	Min(%)	Average(%)	Median(%)	Max(%)
Δ_F^O	5.85	12.16	10.48	35.1
Δ_P^O	0.39	10.06	8.32	48.24
Δ_R^O	0.00	0.66	0.1	7.3

Table 3 summarizes the relative performance of the FCFS, the preemptive-$b\mu/\lambda$ priority and the Myopic($\mathbf{R}$) policies when compared to the optimal policy. The results indicate that the Myopic($\mathbf{R}$) policy yields costs close to the optimal costs.

5 Conclusion and Future Work

In this paper, we consider a system of fleets of machines and assume that each machine is subject to failure. To minimize downtime costs, a spare machine inventory is kept for each fleet. We first address whether these fleets should be served by smaller repair shops dedicated to them or by a centralized repair shop (CRS) serving all fleets. Our numerical study indicates that the CRS system reduces downtime and holding costs significantly. We focus on the CRS systems to determine a practical scheduling policy to reduce the system costs even more. We consider the static FCFS, and preemptive-resume policies from the literature, but more importantly design the new preemptive-Myopic(**R**) policy as an alternative. These three policies are compared with the optimal solution found from the MDP formulation of the problem. Our extensive numerical study demonstrates that the Myopic(**R**) policy outperforms the FCFS, and preemptive-resume priority policies in most cases, and yields costs very close to the optimal.

Our analysis is helpful for OEMs (original equipment manufacturer) that provide maintenance service to their clients. Instead of reserving separate repair crews for each client, they can reduce costs by pooling their personnel and allocate the required number of repair-people to a client whenever the need arises. Additionally, our study demonstrates the benefit of implementing preemptive policies at flexible repair shops that can fix different types of equipment and machinery. If the company has determined that having a certain type of machine functional is more important than another type, our study shows how much cost reduction can be realized by preempting the repair of a less important machinery in favor of the more important one instead following the FCFS policy. In this regard, the Myopic(R) policy is proposed which has straightforward decision rules that can be easily found by even manually computing indices for each class. Our numerical study shows that this policy gives close to optimal costs and usually outperforms a static preemptive-resume policy.

Acknowledgements

This work was supported in part by Natural Sciences and Engineering Research Council (NSERC) of Canada. We would like to thank Dr. Dragan Banjevic who discussed the proof of Proposition 1 with us. We would like to thank the anonymous referee for the invaluable

suggestions which helped us improve the manuscript.

References

Arora, V., F.T.S. Chan, and M.K. Tiwari. 2010. "An integrated approach for logistic and vendor managed inventory in supply chain", *Expert Systems With Applications*, Vol. 37, No. 1, 39–44.

Basten, R.J.I, M.C. van der Heijden, J.M.J. Schutten, and E. Kutanoglu. 2012. "An approximate approach for the joint problem of level of repair analysis and spare parts stocking", *Forthcoming in Annals of Operations Research*, DOI:10.1007/s10479-012-1188-0.

Bitran, G. and R. Caldentey. 2002. "Two-class priority queueing system with state-dependent arrivals", *Queueing Systems*, Vol. 40, 355–82.

Chan, F.T.S. and H. K. Chan. 2004. "Analysis of dynamic control strategies of an FMS under different scenarios", *Robotics and Computer Integrated Manufacturing*, Vol. 20, No. 5, 423–437.

Chan, F.T.S. and A. Prakash. 2012. "Inventory management in a lateral collaborative manufacturing supply chain: A simulation study", *International Journal of Production Research*, Vol. 50, No. 16, 4670–4685.

Chandra, M. J. 1986. "A Study of Multiple Finite-Source Queueing Models", *Journal of the Operational Research Society*, Vol. 37, 275–283.

Chung, S.H., F.T.S. Chan, and H.K. Chan. 2009. "A modified genetic algorithm approach for scheduling of perfect maintenance in distributed production scheduling", *Engineering Applications of Artificial Intelligence*, Vol. 22, No. 7, 1005–1014.

Gross, D. and J. F. Ince. 1981. 'The machine repair problem with heterogenous populations", *Journal of Applied Probability*, Vol. 29, No. 3, 532–549.

Gross, D. and C. M. Harris. 1998. *Fundamentals of Queueing Theory*, John Wiley & Sons, New York.

Gumasta K., F. T. S. Chan, and M.K. Tiwari. 2012. "An incorporated inventory transport system with two types of customers for multiple perishable goods", *International Journal of Production Economics*, Vol. 139, No. 2, 678–686.

Ha, A. 1997. "Optimal Dynamic Scheduling Policy for a Make-to-Stock Production System," *Operations Research*, Vol. 45, 42–53.

Haque, L. and M. J. Armstrong. 2007. "A survey of the machine interference problem", *European Journal of Operational Research*, Vol. 179, No. 2, 469-482.

Iravani, S. M. R., and B. Kolfal. 2005. "When does the $c\mu$ rule apply to finite-population queueing systems?", *Operations Research Letters*, Vol. 33, 301–304.

Iravani, S. M. R., V. Krishnamurthy, and G. H. Chao. 2007. "Optimal server scheduling in nonpreemptive finite-population queueing systems", *Queueing Systems*, Vol. 55, 95-105.

Jaiswal, N. K. 1968. *Priority queues*, Academic Press.

Kelly, F. P. 1975. "Networks of queues with customers of different types", *Journal of Applied Probability*, Vol. 12, 542–554.

Kumar, V., Prakash, M.K. Tiwari, and F.T.S Chan. 2006. "Stochastic make-to-stock inventory deployment problem: an endosymbiotic psychoclonal algorithm based approach", *International Journal of Production Research*, Vol. 44, No. 11, 2245–2263.

Kumar, V., S. Kumar, M.K. Tiwari, and F.T.S. Chan. 2008. "Performance evaluation of flexible manufacturing systems under uncertain and dynamic situations", *Proceedings of The Institution of Mechanical Engineers Part B: Journal of Engineering Manufacture*, Vol. 222, No. B7, 915–934.

Louit, D., R. Pascual, D. Banjevic, and A.K.S. Jardine. 2011. "Optimization models for critical spare parts inventories– a reliability approach", *Journal of the Operational Research Society*, Vol. 62, 992–1004.

Miller, D. R. 1981. "Computation of Steady-State Probabilities for $M/M/1$ Priority Queues", *Operations Research*, Vol. 29, 945–958.

Niño-Mora, J. 2006. "Restless Bandit Marginal Productivity Indices, Diminishing Returns, and Optimal Control of Make-to-Order/Make-to-Stock $M/G/1$ Queues", *Mathematics of Operations Research*, Vol. 31, No. 1, 50–84.

Peña Perez, A., and P. Zipkin. 1997. "Dynamic Scheduling Rules for a Multiproduct Make-to-Stock Queue", *Operations Research*, Vol. 45, No. 6, 919–930.

Sahba, P., B. Balcıoğlu, and D. Banjevic. 2012. "Spare parts provisioning for multiple k-out-of-$n : G$ systems", *Forthcoming IIE Transactions*, DOI:10.1080/0740817X.2012.695102.

Sahba, P. and B. Balcıoğlu. 2011. "The Impact of Transportation Delays on Repairshop Capacity Pooling and Spare Part Inventories", *European Journal of Operational Research*, dx.doi.org/10.1016/j.ejor.2011.05.022.

Sanajian, N., H. Abouee-Mehrizi, and B. Balcıoğlu. 2010. "Scheduling policies in the $M/G/1$ make-to-stock queue", *Journal of the Operational Research Society*, Vol. 61, 115 –123.

Sundarraj, R. P. 2006. "A model for standardizing human decisions concerning service-contracts management", *Annals of Operations Research,* Vol. 143, 171–189.

Sztrik, J. 2001. "Finite source queueing systems and their applications: A bibliography", *Research Report, Institute of Mathematics and Informatics, University of Debrecen, Debrecen, Hungary.*

Taylor, J. and R. R. P. Jackson. 1954. "An application of the birth and death process to the provision of spare machines", *Operational Research Quarterly*, Vol. 5, 95–108.

Tijms, H. C. 2003. *A First Course in Stochastic Models*, John Wiley & Sons Ltd, West Sussex, England.

Veatch, M., and L. M. Wein. 1996."Scheduling a make-to-stock queue: Index policies and hedging points", *Operations Research*, Vol. 44, No. 4, 634–647.

Veran, M. 1984. "Exact analysis of a priority queue with finite source", *Proceedings of the*

International Seminar on Modelling and Performance Evaluation Methodology, Paris.

de Véricourt, F., F. Karaesmen, and Y. Dallery. 2000. "Dynamic Scheduling in a Make-to-Stock System: A Partial Characterization of Optimal Policies", *Operations Research*, Vol. 48, No. 5, 811–819.

Wadhwa, S., Bibhushan, and F.T.S. Chan. 2009. "Inventory performance of some supply chain inventory policies under impulse demands", *International Journal of Production Research*, Vol. 47, No. 12, 3307–3332.

Waters, D. 2003. *Logistics: An Introduction to Supply Chain Management*, Palgrave Macmillan, New York.

Wein, L. M. 1992. "Dynamic Scheduling of a Multiclass Make-to-Stock Queue", *Operations Research*, Vol. 40, No. 4, 724–735.

Wong, C.S., F.T.S. Chan, and S.H. Chung. 2012. "A genetic algorithm approach for production scheduling with mould maintenance consideration", *International Journal of Production Research*, Vol. 50, No. 20, 5683–5697.

Yu, Y., S. Benjaafar, and Y. Gerchak. 2009. "Capacity sharing and cost allocation among independent firms in the presence of congestion", *under review.*

Zorna, W. L., R. F. Deckroa, and L. J. Lehmkuhlb. 1999. "Modeling diminishing marginal returns in a hierarchical inventory system of reparable spare parts", *Annals of Operations Research*, Vol. 91, 319–337.

Zheng, Y., and P. Zipkin. 1990. "A queueing model to analyze the value of centralized inventory information", *Operations Research*, Vol. 38, No. 2, 296–307.

Zipkin, P. 1995. "Performance analysis of a multi-item production-inventory system under alternative policies," *Management Science*, Vol. 41, No. 4, 690–703.