

# Policies for a Spare Parts Provisioning Problem

**Pedram Sahba, Barış Balcıoğlu, Dragan Banjević**

University of Toronto, Department of Mechanical and Industrial Engineering  
5 King's College Rd., Toronto, ON M5S 3G8, CANADA,  
*pedram@mie.utoronto.ca, baris@mie.utoronto.ca, banjev@mie.utoronto.ca*

## Abstract

In this paper, we analyze a system with different fleets of machines and assume that each machine can fail due to a single type of repairable critical component. The system aims to have all machines in each fleet be functional at all times to continue production at targeted levels. Therefore, a certain number of critical components are kept in a centralized inventory as spare parts. Failed components are repaired in a single repair shop. We consider three types of pooled inventory structure: (i) a shared inventory serving all fleets, (ii) reserved inventories for each fleet, (iii) and a mixture of the two. Additionally, we consider alternative dispatching rules for the repaired component. The destination fleet for a repaired component can be chosen either on a first-come-first-served basis or by considering static or dynamic priority rules among the fleets. The alternative inventory structures and dispatching rules result in 7 alternative policies. Our analysis indicates that a system with a mixture of a shared inventory and possibly reserved inventories for each fleet together with a static-priority dispatching rule is the best policy in terms of minimizing the cost. Our numerical examples indicate that this policy is also equivalent to a special type of a system operating under the inventory rationing policy. However, when transportation times are considered, the hybrid policies surpass the performance of the inventory rationing policy.

**Keywords and Phrases:** Spare parts, Multiple finite-population queueing systems, Priority queues, Hybrid policies, Inventory rationing

# 1 Introduction

For many companies the recent economic downturn has increased the need to extend the lifetime of their existing equipment for production. To achieve this goal, maintenance jobs are more frequently outsourced to companies such as ABB and Advanced Technology Services (Economist<sup>1</sup>). Among other things, the outsourcing companies consider consolidating spare part suppliers to reduce asset management costs for their clients. Consolidation can be in the form of pooling repair shop capacity and spare part inventories. When equipment stoppages arise due to critical component failures, repairing the failed components can lower the cost. This can also be a more environmentally friendly practice than using non-repairable components. To increase the availability of the production system, keeping spare parts of critical repairable components should also be carefully considered. In this study, we analyze a problem where machines in different fleets are subject to failure due to a single critical component. We consider different pooled inventory topologies and various dispatching rules from a single repair shop where the failed components from all fleets are fixed. After describing each policy, we compare them to address three questions: (i) Should there be a separate repair shop (plus a separate inventory) for each fleet or should a single repair shop with a higher capacity serve all the fleets? (ii) Given a shared high capacity repair shop, should we reserve a separate spare parts inventory for each fleet, or should a single spare parts inventory be shared by all fleets, or would a mixture of the two be more cost efficient? (iii) And finally, when a repaired component from the repair shop is dispatched, should we choose the destination fleet/reserved inventory according to a first-come-first-served (FCFS) rule or could static or dynamic rules of prioritization of fleets make this decision more cost effective? The answers to these questions are important for the outsourcing companies that aim to minimize their system costs.

We will develop queueing based solutions to address these questions. Our problem is an example of a Markovian queueing system with multiple finite calling populations (see Sztrik,

---

<sup>1</sup>“Outsourcing: A quick fix”, Vol. 390, No: 8617, February 7-13th, 2009, p. 58

2001, for a comprehensive bibliography on systems with finite populations). The model in which a separate repair shop and inventory are allocated for each fleet can be easily analyzed using a birth-and-death model, e.g., Taylor and Jackson (1954). However, including a joint repair shop is difficult even when an FCFS dispatching rule is followed. The earlier work in the literature on machine interference or machine repairperson problems addresses this without considering inventories. Chandra (1986) employs mean value analysis as the only suitable analytical/numerical technique for the FCFS repair policy for  $m$  fleets of machines sharing a single repair shop. He models the non-preemptive priority policy in the same study, too.

When priorities are also introduced in finite population systems, the analysis becomes more challenging. Miller (1981) presents recursive computational formulae to obtain the steady-state distributions of customers in a two-priority (preemptive and non-preemptive) class Markovian single server queue. Veran (1984) analyzes the same system assuming a preemptive-resume policy and avoids the computational complexity of the method due to Jaiswal (1968). Bitran and Caldentey (2002) investigate a two-priority class queueing system with state-dependent arrival rates operating under the preemptive-resume priority policy. They present a general approach for computing the steady-state distribution of the number of customers in the system for each class. Iravani, Krishnamurthy, and Chao (2007) consider a Markovian finite-population queueing system with heterogeneous fleets of machines repaired by a single server. They prove that when preemption is not allowed, the optimal policy is a simple static non-preemptive priority policy and they present the sufficient conditions to prioritize the classes correctly. Iravani and Kolfal (2005) study the same problem when preemption is permitted and show under which conditions a static preemptive-resume priority policy is optimal.

Our first contributions in this paper are to analyze a Markovian single server queueing system with multiple classes of customers whose arrival rates are state-dependent. Under the FCFS policy, when the exponential service/repair rate is the same for all classes, we obtain the exact steady-state distribution of the number of customers in the system for each

class as an alternative to the mean-value analysis. Then, we extend the model due to Bitran and Caldentey to more than 2 priority classes of customers, in which different exponential repair rates can also be assumed for each class. An immediate benefit of this extension is that when the optimal repair policy in a system without inventories is the static preemptive-resume policy, the cost of the system studied by Iravani and Kolfal with  $m$  fleets of machines can be now computed.

More importantly, these models introduce spare part inventories to our problem, which considers multiple finite population customer classes. This is important because earlier studies with spare part inventories usually assume a single finite population. Graves and Keilson (1983) consider a fleet of airplanes each of which requires one unit from a set of different types of components be available. Under budgetary constraints, they determine the optimal stocking levels for each component with the objective of maximizing system availability defined as each component inventory being above a critical level. Dshalalow (1991) models a system with  $N$  identical machines subject to failure and  $S$  spare machines. When a machine fails, it is immediately replaced by a spare machine if there is one. In this system, by permitting non-Markovian failure and repair processes, he obtains the steady-state distribution of the number of jobs in the repair shop. Abboud and Daigle (1997) consider a fleet of machines some of which are operational and the others kept as spare machines. The failed machine is sent to a multi-server repair shop, but to be repaired, it requires that a spare critical component be available. If such a critical component is not available, it has to be ordered and it is received at the end of an exponentially distributed lead time. Upon completion of the repair, the machine joins the spare machine group if demand is met by the available functional machines. Otherwise, the repaired machine is immediately deployed.

In our problem, we consider different fleets of machines. Ordering new critical components is excluded from the analysis since our models determine the optimal number of repairable spare parts; if the system has no spare components when a machine fails, the machine is down until the repair shop can dispatch a repaired component. The alternative

policies we model are inspired from the production/inventory systems modeled as make-to-stock queueing systems. In the production/inventory setting, a single server models the production stage (corresponding to our repair shop) and inventories of finished products are kept (corresponding to our spare part inventories) to minimize backordering or lost sales costs (corresponding to our down time costs). The fundamental difference between our problem and production/inventory systems is that the rates of the demand processes are constant whereas in our case the number of operational machines determines the failure rate, thus, the arrival rate of components to the repair shop. Although, at first glance, this appears to be a minor distinction, as will be discussed further, the state-dependent arrival rates make our problem difficult to analyze.

Therefore, we will next summarize certain studies from the production/inventory literature. In all these papers, the transportation times and costs between production facility and inventories are ignored. Benjaafar, Cooper, and Kim (2005) model a production system satisfying demand from  $m$  classes with an FCFS dispatching policy. The two alternatives that they compare are what we will refer to as the RIF (reserved inventory-FCFS) system (or policy to be used interchangeably) where a separate inventory is reserved/allocated for each class and the SIF (shared inventory-FCFS) system where a shared inventory is held for all classes. They show that if backordering costs are the same for all classes, the SIF system gives the same cost as or less cost than the RIF system. This conclusion is in agreement with Eppen (1979) who first showed the benefits of holding a single inventory for all classes instead of keeping separate inventories. In another system with reserved inventories, different classes are given different priorities, what we will call the RIP (reserved inventory-priority) system. Sanajian, Abouee-Mehrizi, and Balcioğlu (2010) study this system in the  $M/G/1$  queueing setting where the finished product is sent to the highest-priority class if its inventory level is below its base-stock level.

The RIP policy is not the only applicable policy in systems if different classes have different priorities. As an alternative, inventory rationing (IR) policy can be used. While serving  $m$  classes, the IR policy uses threshold inventory levels  $R_r$ ,  $r = 1, \dots, m + 1$  with

$R_1 = 0$  and  $R_{m+1} = S$  ( $S$  is the base-stock level and when the inventory level is  $S$  the production stops). If the stock level is between  $R_{r+1}$  and  $R_r$  (with  $R_{r+1} > R_r$ ), only demand requests of classes 1 to  $r$  are satisfied on an FCFS basis, and demand from other classes is backlogged or lost. Thus, if the inventory level is between  $R_{r+1}$  and  $R_r$  even if there are pending orders from classes  $r + 1$  to  $n$ , the completed product is placed in inventory. When there is no positive-stock, the finished product is sent out to the highest-priority customer backlogged if backlogging is permitted. Ha (1997a) models a Markovian multi-class single server system with a centralized inventory in which unsatisfied demands are lost. Ha (1997b) studies the same problem with two classes of customers when backordering is allowed. In both studies, he proves that in systems with centralized inventories, the IR policy is the optimal production control policy. de Véricourt, Karaesmen, and Dallery (2001) prove that the IR policy is optimal when serving  $m$  classes of customers from a pooled inventory when backordering is allowed. de Véricourt, Karaesmen, and Dallery (2002) provide an algorithm to compute the optimal rationing levels and the cost of the IR policy in  $M/M/1$  systems serving  $m$  classes of customers when backordering is allowed. Kranenburg and van Houtum (2008) consider the inventory rationing policy for the spare part provisioning problem by assuming constant failure rates for non-repairable components. de Véricourt, Karaesmen, and Dallery (2001) also introduce the strict priority (SP) policy, under which demand is satisfied on an FCFS basis as long as there is stock on hand. If the SP system has backorders, the finished product is dispatched to the highest-priority class among those that are backordered. Abouee-Mehrizi, Balcioglu and Baron (2010) obtain the optimal costs for the SP and IR policies in  $M/G/1$  systems.

In this paper, we consider a repair shop/spare part inventory system where the SIF, RIF, RIP, SP policies and the IR policy with 3 threshold levels can be implemented when serving  $m$  fleets of machines. However, the state-dependent arrivals at the repair shop prevent us from proving certain conclusions derived in production/inventory systems. For instance, while comparing the SIF and RIF systems, Benjaafar, Cooper, and Kim can assume that the order arrival and the service completion instants in one system coincide exactly with

those in the other system. This is because the arrivals are not affected by the inventory levels and do not depend on whether or not there are reserved inventories. Using this, they can show that the SIF system results in a cost less than or equal to that of the RIF system when the holding cost rate is the same in both systems and the backordering cost rate is the same for each class. In our problem, on the other hand, the arrival rates depend on the inventory levels and which policy is used. Therefore, a similar path-by-path basis comparison is not possible. Similarly, both the SP and SIF production/inventory systems, when they are in the backordering period, will have the same total number of backorders given the same arrivals. The SP and SIF systems will differ in the ratios of the expected type  $i$  backorders to the expected total number of backorders. Using this, as in de Véricourt, Karaesmen, and Dallery (2001) for  $M/M/1$  systems or in Abouee-Mehrzi, Balcioglu and Baron for  $M/G/1$  systems, it is possible to show that in terms of minimizing the system cost, the SP system is as good as or better than the SIF system. When it comes to our problem, the SP and SIF systems no longer have the same total number of down machines, which prevents us from proving the same relation between the two systems in our setting.

However, we make several other contributions by extending the application of these policies to our problem. We also introduce two more policies. First, the HP (hybrid priority) system has a shared inventory for all fleets and may have reserved inventories for some fleets. The shared inventory is depleted on an FCFS basis, and when it is empty, the repair shop dispatches repaired components to fleets/reserved inventories considering their priorities. Second, the HF (hybrid FCFS) system is similar to the HP system except that when its shared inventory is depleted, the repair shop dispatches repaired components to the fleet (or the inventory reserved for that fleet) with the longest waiting repair order.

From our total of 7 policies, we establish formal relations in terms of the system cost between certain policies. Since it is not possible to rank all these policies analytically, in Section 5, we present the results of an extensive numerical study. We can summarize our conclusions in the following way. We start with the system in which a separate repair shop and inventory are allocated for each fleet. We refer to this system as the base case (BC)

system. We add the repair rates of the separate repair shops and use the total as the repair rate of the single repair shop of the 7 policies. This is a common technique for modeling resource capacity pooling (see Yu, Benjaafar, and Gerchak, 2009, and the references therein for capacity pooling in the production/inventory systems). Our numerical results show that all 7 policies with a pooled repair shop capacity are superior to the BC system in terms of reducing system costs. Next, we compare the relative performances of the 7 policies. Our results indicate that the HP and the IR systems turn out to be equivalent and are both superior to the remaining 5 policies in reducing system costs.

In this paper, similar to the literature on production/inventory systems modeled by make-to-stock queues we ignore transportation times of spares and repaired components to the fleets and failed components to the repair shop. Incorporating transportation times in analytical models is quite difficult. For instance, for a single fleet, Gross, Kahn and Marsh (1977) relax the assumption of state-dependent failure rates to a constant failure rate and use ample server queueing systems to incorporate transportation times in their analysis. However, the effects of transportation delays on system performance can be significant. For instance, Sahba and Balcioglu (2010) compare the BC system and the RIF system and using the mean value analysis show that when the mean transportation times between fleets and repair shop exceed the mean repair time, repair shop pooling may not be beneficial anymore. Therefore, in Section 5, using a simulated example we also test the impact of non-negligible transportation times on the performances of the HF, HP and IR policies. The results indicate that the HP policy is the clear cut winner and the costs under the IR policy steeply increase with longer transportation times compared to modest increases under the HF and HP policies.

The rest of the paper is organized as follows. In Section 2, we present our problem and the alternative systems we study. In Section 3, we analyze the FCFS based systems, namely the RIF, HF and SIF systems. In Section 4, the RIP, HP, SP and IR systems, i.e., systems in which different fleets are given different priorities, are modeled. In Section 5, we present numerical results comparing the relative performances of all of these policies.

## 2 Alternative Policies

We consider an outsourcing company that serves  $m$  fleets of machines parameterized by  $i = 1, \dots, m$ . Each fleet  $i$  consists of  $N_i$  machines (interchangeably can be referred to as type  $i$  machines). The goal is to have all machines in each fleet be functional at all times to continue production at targeted levels. However, this might not be possible since each machine is subject to failure due to a single type of repairable critical component. Times to failure, that is the periods between installation of a new or repaired component on a machine and the next failure instant, follow an exponential distribution with rate  $\lambda_i$  for type  $i$  machines (implying that each repair makes the component as good as new, and the failure rate only depends on which fleet will use it next). Different failure rates can be due to different machine types from one fleet to another that use the same critical component or the same type of machines in different fleets operated under different environmental conditions. When a component fails, it is sent to a repair shop, which is modeled as an FCFS single server queue where the repair times are independent and identically distributed (i.i.d.) exponential random variables (r.v.s) with rate  $\mu$ . If there is a stock of critical components kept as spare parts, one can install a spare component immediately on the failed machine to prevent production loss. Otherwise, the failed machine is down until a repaired component can be dispatched from the repair shop. During such down times, the outsourcing company is penalized and incurs a backordering/down time cost of  $b_i$  per type  $i$  machine down per unit time. In other words, keeping spare part inventories might help decrease the backordering cost at the expense of incurring inventory holding cost. The outsourcing company incurs a holding cost of  $h$  per spare component kept in the inventory per unit time. In this problem, the outsourcing company can keep a shared spare part inventory for all fleets, or reserve separate inventories for each fleet, or consider a mixture of these two, which will be referred to as the *hybrid* model. We assume that the transportation time between fleets and the repair shop plus pooled inventory, times to install and remove a component, and transportation costs are negligible. We will consider non-negligible transportation times in Section 5.4.

The system presented above involves a component allocation problem and can be managed using different policies. These policies depend on the inventory structure and whether different fleets are prioritized and whether dispatching decisions are made at the beginning of the repair service or are postponed until the end of the repair service. In this paper, for non-FCFS policies, we focus on postponement of the allocation since it often reduces costs compared to not postponing. Accordingly, the alternative policies we will consider are:

**RIF Policy** The *reserved inventory-FCFS* system: The outsourcing company reserves a separate inventory for each fleet operating according to a base-stock policy. The repair shop dispatches the repaired component in an FCFS manner. We study this policy in Section 3.1 where we obtain the optimal base-stock levels and the optimal system cost.

**HF Policy** The *hybrid FCFS* system: In addition to a reserved inventory for each fleet, there is a shared inventory for all fleets, and each inventory operates according to a base-stock policy. The repair shop dispatches the repaired component in an FCFS manner. The outsourcing company first expends spare parts from the shared inventory and only when it is depleted, the reserved inventories are used. We study this policy in Section 3.2 where we obtain the optimal base-stock levels and the optimal system cost.

**SIF Policy** The *shared inventory-FCFS* system: There is a shared inventory for all fleets operating according to a base-stock policy. The repair shop dispatches the spare part or the repaired component in an FCFS manner. We study this policy in Section 3.3 where we obtain the optimal base-stock levels and the optimal system cost.

**RIP Policy** The *reserved inventory-priority* system: The outsourcing company reserves a separate inventory for each fleet operating according to a base-stock policy. Moreover, fleets are ranked according to a priority scheme. The repair shop dispatches the repaired component to the fleet with the highest priority among those that have pending orders. We study this policy in Section 4.1 where we obtain the optimal base-stock levels and the optimal system cost.

**HP Policy** The *hybrid priority* system: In addition to a reserved inventory for each fleet, there is an inventory shared by all fleets, and each inventory operates according to a base-stock policy. The outsourcing company first dispatches spare parts from the shared inventory on an FCFS basis and only when it is depleted, are the reserved inventories used. When there are no spare parts in the shared inventory, the repaired component is used to serve the highest priority fleet among those with pending orders. We study this policy in Section 4.2 where we obtain the optimal base-stock levels and the optimal system cost.

**SP Policy** The *strict priority* system: There is a shared inventory for all fleets operating according to a base-stock policy. As long as there is stock in the inventory, spare parts are dispatched on an FCFS basis. When there are backlogs, the repaired component is dispatched to the fleet that has the highest priority among those with pending orders in the system. We study this policy in Section 4.3 where we obtain the optimal base-stock level and the optimal system cost.

**IR Policy** The *inventory rationing* system with three rationing levels: There are three threshold inventory levels  $R_3 \geq R_2 \geq R_1 = 0$ , and fleets 1 to  $m$  are prioritized from highest to lowest. If the inventory level is between  $R_3$  and  $R_2$ , failed components are replaced from the inventory on an FCFS basis. If the inventory level is between  $R_1$  and  $R_2$ , a) only failed components from fleet 1 are replaced from the inventory whereas the others are backlogged, and b) each repaired component is placed in the inventory. Only when the inventory level is  $R_2$ , is a repaired component dispatched to the highest priority fleet among those with down machines. We study this policy in Section 4.4 where we obtain the optimal rationing levels and the optimal system cost.

### 3 Systems with FCFS Dispatching Policies

In this section, we will analyze three systems with FCFS dispatching policy. We will start with the RIF policy in Section 3.1 since its results are necessary for analyzing the HF policy in Section 3.2. Finally, the SIF policy as a special case of the HF policy will be presented in Section 3.3. The model in Section 3.1 can be employed to obtain the distribution of the number of type  $i$  customers in Markovian systems where different classes have state-dependent arrival rates and are served by a single server with the same exponential service rate for all customers.

#### 3.1 The RIF Policy: The FCFS System with Reserved Inventories

In this section, we study the system in which an inventory of  $S_i \geq 0$  spare parts is reserved for fleet  $i$ /type  $i$  machines,  $i = 1, \dots, m$ . When a type  $i$  machine fails, if there is positive stock in its inventory, one of the spare parts is installed on that machine, thereby avoiding any machine stoppages. If there is no positive stock for type  $i$  machines when the critical component fails, the machine is down until a repaired component can be sent from the repair shop. Thus, we exclude the possibility of using a spare part from the positive inventory reserved for another fleet (permitting this would make this system the SIF system studied in Section 3.3). A type  $i$  machine failure results in a type  $i$  repair order, which is queued to be serviced in an FCFS single server repair shop. When a type  $i$  repair is completed, the repaired component is installed on a type  $i$  machine that is down if there are any such down machines. Otherwise, it is placed in the inventory of fleet  $i$ , raising the inventory level by 1. Let  $O_i(t)$  be the number of type  $i$  orders in the repair shop at time  $t$ . Then, we have the following relation between the number of type  $i$  orders in the repair shop and the inventory and backorder levels of type  $i$  machines: If  $O_i(t) \leq S_i$ , the inventory level for type  $i$  machines is  $I_i(t) = S_i - O_i(t)$  and all  $N_i$  machines are operational, and if  $S_i < O_i(t) \leq N_i + S_i$ , fleet  $i$  has  $O_i(t) - S_i$  down machines, i.e., components backordered.

Our objective is to determine the optimal  $S_i^*$  levels to reserve for each fleet  $i$  to minimize

the long-run average system cost as a time-average. Let  $p^i(k) := P(O_i = k)$  be the steady-state probability of having  $k$  type  $i$  orders in the repair shop. The objective function can be expressed as follows:

$$C_{RIF}^* \equiv \min_{\mathbf{S}} \{C(\mathbf{S}) : C(\mathbf{S}) = \sum_{i=1}^m C_i(S_i)\}, \quad (1)$$

where  $\mathbf{S} = (S_1, S_2, \dots, S_m)$  and  $C_i(S_i)$  is the cost due to fleet  $i$  given by

$$C_i(S_i) = h \sum_{k=0}^{S_i} (S_i - k)p^i(k) + b_i \sum_{k=S_i+1}^{N_i+S_i} (k - S_i)p^i(k). \quad (2)$$

In order to obtain  $p^i(k)$ , we start by characterizing the system state at an arbitrary time by the vector  $(x_1, x_2, \dots, x_n)$  stating that there are a total of  $n$  repair orders where  $x_j = i$  means that the  $j$ th order from the end of the repair queue is type  $i$ , i.e., originating from fleet  $i$ . Hence,  $x_n$  and  $x_1$  are the first and last repair orders in the system, respectively,  $x_n$  being the one under repair. Whenever it is not necessary to present the entries of the state vector, we will use  $\omega_{\mathbf{n}}$  as a shorthand notation for  $(x_1, x_2, \dots, x_n)$ .

We define  $Y_i(\omega_{\mathbf{n}}) = \sum_{j=1}^n I(x_j = i)$  as the number of type  $i$  orders, where  $I(E)$  is the indicator function, which equals 1 if event  $E$  is true and 0 otherwise. Accordingly,  $Y_i(\omega_{\mathbf{n}}) \in \{0, 1, \dots, N_i + S_i\}$ . Observe that the failure rate from fleet  $i$  depends only on  $Y_i(\omega_{\mathbf{n}})$  (equivalently, how many type  $i$  machines are operational). We define this failure rate  $\tilde{\Lambda}_i(\omega_{\mathbf{n}})$  as

$$\begin{aligned} \tilde{\Lambda}_i(\omega_{\mathbf{n}}) &= \begin{cases} N_i \lambda_i, & \text{if } 0 \leq Y_i(\omega_{\mathbf{n}}) \leq S_i, \\ (N_i + S_i - Y_i(\omega_{\mathbf{n}})) \lambda_i, & \text{if } S_i < Y_i(\omega_{\mathbf{n}}) \leq N_i + S_i, \\ 0, & \text{otherwise} \end{cases} \\ &= (N_i + S_i - \max\{Y_i(\omega_{\mathbf{n}}), S_i\}) \lambda_i. \end{aligned} \quad (3)$$

Adding the failure rates from all fleets,  $\tilde{\Lambda}(\omega_{\mathbf{n}}) = \sum_{i=1}^m \tilde{\Lambda}_i(\omega_{\mathbf{n}})$  gives us the state-dependent arrival rate of repair orders to the single server queue modeling the repair shop facility.

Let  $p_n(\omega_{\mathbf{n}})$  be the steady-state probability of being in state  $\omega_{\mathbf{n}}$ . We will now relate probabilities of interest to one another. Let  $p_0(0)$  denote the probability that the repair shop

is idle and  $N_i$  machines are running and  $S_i$  spare parts are available in the inventory for each fleet  $i$ . We can write the global balance equations:

$$-p_0(0)\Lambda + [p_1(1) + \dots + p_1(m)]\mu = 0, \quad (4)$$

where  $\Lambda = \sum_{i=1}^m N_i \lambda_i$  and for a feasible  $x_1 \in \{1, \dots, m\}$  and setting  $N = \sum_{i=1}^m (N_i + S_i)$  (the maximum number of components that can be in the repair shop),

$$p_{n-1}(x_2, \dots, x_n) \tilde{\Lambda}_{x_1}(x_2, \dots, x_n) - p_n(x_1, x_2, \dots, x_n) [\mu + \tilde{\Lambda}(x_1, x_2, \dots, x_n)] \\ + [p_{n+1}(x_1, x_2, \dots, x_n, 1) + \dots + p_{n+1}(x_1, x_2, \dots, x_n, m)]\mu = 0, \quad 1 < n \leq N - 1, \quad (5)$$

and finally,

$$p_{N-1}(x_2, \dots, x_N) \tilde{\Lambda}_{x_1}(x_2, \dots, x_N) - p_N(x_1, x_2, \dots, x_N) [\mu + \tilde{\Lambda}(x_1, x_2, \dots, x_N)] = 0. \quad (6)$$

Note that some states might be infeasible, in which case the corresponding probabilities in Eq.s (5-6) are zero.

We assume that  $p_N(\omega_{\mathbf{N}}) = p_N(\omega'_{\mathbf{N}}) = p_N > 0$  for any  $\omega_{\mathbf{N}}$  and  $\omega'_{\mathbf{N}}$ . We will obtain the limiting probabilities expressing all probabilities in terms of  $p_N$  and using a normalization constraint. Since the underlying Markov chain is irreducible and has a finite number of states, it is necessarily positive recurrent. Therefore, the limiting distribution to be found is unique (Bhat and Miller, 2002, p. 222) implying that the solution we will provide is the only possible solution.

To obtain the limiting probabilities, we will start from Eq. (6). Note that  $\tilde{\Lambda}(x_1, \dots, x_N) = 0$  since there are no machines operational. In Eq. (6),  $p_N(x_1, x_2, \dots, x_N) = p_N$  and  $\tilde{\Lambda}_{x_1}(x_2, \dots, x_N) = \lambda_i$  for  $x_1 = i \in \{1, \dots, m\}$  because there is only one type  $i$  machine functional, which can fail. In state with  $N$  components in the repair shop, if we fix  $x_1 = i$  for some  $i$ ,  $x_1$  with each ordering of  $x_2, \dots, x_N$  is a different  $\omega_{\mathbf{N}}$ , and the probability of being in any one of them is the same,  $p_N$ . For each such ordering  $\omega_{\mathbf{N}}$ , dropping  $x_1 = i$  gives the  $\omega_{\mathbf{N}-1}$  that satisfies Eq. (6), and  $p_{N-1}(\omega_{\mathbf{N}-1}) = \mu p_N / \lambda_i$ . Thus,  $\omega_{\mathbf{N}-1}$  is a generic state with only one machine of type  $i$  functional, and the probability of being in that state does not

depend on the sequence of repair orders but on which type of machine is the only operational machine (type  $i$  machine in this example).

Next, after rearranging Eq. (5) and substituting  $N - 1$  for  $n$ , we obtain

$$\begin{aligned}
p_{N-2}(x_2, \dots, x_{N-1})\tilde{\Lambda}_{x_1}(x_2, \dots, x_{N-1}) &= \mu p_{N-1}(x_1, \dots, x_{N-1}) \\
&+ p_{N-1}(x_1, \dots, x_{N-1})\tilde{\Lambda}(x_1, \dots, x_{N-1}) \\
&- \mu p_N(x_1, \dots, x_{N-1}, x_N). \tag{7}
\end{aligned}$$

Observe that in state  $(x_1, \dots, x_{N-1})$ , there is one machine (say type  $i$ ) operational, which can fail. This forces  $\tilde{\Lambda}(x_1, \dots, x_{N-1}) = \lambda_i$  and  $p_N(x_1, \dots, x_{N-1}, x_N = i)$ . Given our assumption, we have  $p_N(x_1, \dots, x_{N-1}, i) = p_N(i, x_1, \dots, x_{N-1})$ . From earlier discussion making use of Eq. (6), we showed that  $\lambda_i p_{N-1}(x_1, \dots, x_{N-1}) = \mu p_N(i, x_1, \dots, x_{N-1})$ , which implies  $\lambda_i p_{N-1}(x_1, \dots, x_{N-1}) = \mu p_N(x_1, \dots, x_{N-1}, i)$ . This helps the last two terms on the RHS of Eq. (7) cancel out. In state  $N - 1$  with one type  $i$  machine functional, if we fix  $x_1 = j$ ,  $x_1$  with each ordering of  $x_2, \dots, x_{N-1}$  is a different  $\omega_{\mathbf{N}-1}$  and the probability of being in any one of them is the same,  $\mu p_N / \lambda_i$ . For each such ordering  $\omega_{\mathbf{N}-1}$ , dropping  $x_1 = j$  gives the  $\omega_{\mathbf{N}-2}$  with one type  $i$  and one type  $j$  machines functional ( $i$  and  $j$  can be the same) satisfying Eq. (7) as  $p_{N-2}(\omega_{\mathbf{N}-2})\Lambda_j(\omega_{\mathbf{N}-2}) = \mu p_{N-1}(\omega_{\mathbf{N}-1}) = \mu^2 p_N / \lambda_i$ . This again brings us to the conclusion that  $p_{N-2}(\omega_{\mathbf{N}-2})$  is (just like  $\tilde{\Lambda}_j(\omega_{\mathbf{N}-2})$ ) independent of how different types of repair orders are sequenced and depending only on the types of the two operational machines.

For the general case ( $1 \leq n \leq N - 3$ ), we are going to prove by induction that  $p_n(\omega_{\mathbf{n}})$  depends only on the number of failed components from each fleet (equivalently, the number of operational machines in each fleet). Assume that  $p_{n+2}(\omega_{\mathbf{n}+2})$  and  $p_{n+1}(\omega_{\mathbf{n}+1})$  depend only on the number of failed components from each fleet. This implies  $p_{n+2}(x_1, x_2, \dots, x_{n+1}, i) = p_{n+2}(i, x_1, x_2, \dots, x_{n+1})$  and  $p_{n+1}(x_1, x_2, \dots, x_{n+1})\tilde{\Lambda}_i(x_1, \dots, x_{n+1}) = \mu p_{n+2}(i, x_1, x_2, \dots, x_{n+1})$ . Summing up over all  $i$ , we have

$$p_{n+1}(x_1, x_2, \dots, x_{n+1})\tilde{\Lambda}(x_1, \dots, x_{n+1}) = \mu \sum_{i=1}^m p_{n+2}(x_1, x_2, \dots, x_{n+1}, i).$$

Using this equality above in Eq. (5) results in cancelations, and we arrive at

$$p_n(x_2, \dots, x_{n+1})\tilde{\Lambda}_{x_1}(x_2, \dots, x_{n+1}) = \mu p_{n+1}(x_1, \dots, x_{n+1}).$$

In other words, given that the state  $\omega_{\mathbf{n}}$  has one less failed component of type  $x_1$  than the state  $\omega_{\mathbf{n}+1}$ , we have established  $p_n(\omega_{\mathbf{n}})\tilde{\Lambda}_{x_1}(\omega_{\mathbf{n}}) = \mu p_{n+1}(\omega_{\mathbf{n}+1})$ . Thus,  $p_n(\omega_{\mathbf{n}})$  also depends only on the number of failed components from each fleet, not how they are ordered in the repair shop queue.

Let us consider a state  $\omega_{\mathbf{n}}^j$  with  $n$  repair orders such that there are  $y_1$  components from fleet 1,  $y_2$  from fleet 2, ..., and  $y_m$  from fleet  $m$  at the repair shop, i.e.,  $y_1 = Y_1(\omega_{\mathbf{n}}^j), \dots, y_m = Y_m(\omega_{\mathbf{n}}^j)$  and  $y_1 + y_2 + \dots + y_m = n$ . We use a superscript  $j$  because there are  $K$  such states differing from one another due to the ordering of repair orders, and  $p_n(\omega_{\mathbf{n}}^j) = q(y_1, \dots, y_m)$  is the same for all  $j$ ,  $j = 1, \dots, K$ , where

$$K = \binom{n}{y_1, \dots, y_m}.$$

Recall that  $\tilde{\Lambda}_i(\omega_{\mathbf{n}}^j)$  only depends on  $y_i$ , which makes  $\tilde{\Lambda}(\omega_{\mathbf{n}}^j) = \sum_{i=1}^m \tilde{\Lambda}_i(\omega_{\mathbf{n}}^j)$  dependant only on  $(y_1, \dots, y_m)$ . Then in the remainder of the discussion, we can make use of  $\Lambda_i(y_i)$ , the failure rate of fleet  $i$  when there are  $y_i$  type  $i$  orders in the repair shop. Our discussion so far has showed that  $q(y_1, \dots, y_i, \dots, y_m)\Lambda_i(y_i) = \mu q(y_1, \dots, y_i + 1, \dots, y_m)$ . Summing up over all  $i$ , we obtain

$$q(y_1, \dots, y_m) = \frac{\mu}{\Lambda(y_1, \dots, y_m)}(q(y_1+1, y_2, \dots, y_m) + q(y_1, y_2+1, \dots, y_m) + \dots + q(y_1, y_2, \dots, y_m+1)), \quad (8)$$

where due to Eq. (3),  $\Lambda(y_1, \dots, y_m) = \sum_{i=1}^m \Lambda_i(y_i) = \sum_{i=1}^m (N_i + S_i - \max\{y_i, S_i\})\lambda_i$  as the total failure rate given  $(y_1, \dots, y_m)$ . However, we are interested in the probability of being in any of these  $j$  states with  $y_1$  components from fleet 1,  $y_2$  from fleet 2, ..., and  $y_m$  from fleet  $m$ , which we will denote by  $p(y_1, y_2, \dots, y_m)$ :

$$p(y_1, y_2, \dots, y_m) = \sum_{j=1}^K p_n(\omega_{\mathbf{n}}^j) = Kq(y_1, \dots, y_m) = q(y_1, \dots, y_m) \binom{n}{y_1, y_2, \dots, y_m}.$$

If we multiply both sides of Eq. (8) by  $K$ , we arrive at

$$\begin{aligned}
p(y_1, y_2, \dots, y_m) &= \frac{\mu}{(n+1)\Lambda(y_1, \dots, y_m)} \{ (y_1+1)p(y_1+1, y_2, \dots, y_m) \\
&\quad + (y_2+1)p(y_1, y_2+1, \dots, y_m) + \dots \\
&\quad + (y_m+1)p(y_1, y_2, \dots, y_m+1) \}. \tag{9}
\end{aligned}$$

Using Eq. (9), we can express all  $p(y_1, y_2, \dots, y_m)$  in terms of  $p_N$ , which is the probability of having  $N$  repair orders in the system. After employing the normalization constraint,

$$\sum_{n=0}^N \sum_{y_1+\dots+y_m=n} p(y_1, y_2, \dots, y_m) = 1,$$

we obtain  $p_N$  and all  $p(y_1, y_2, \dots, y_m)$  where  $y_i \in \{0, 1, \dots, N_i + S_i\}$  for  $i = 1, \dots, m$ .

Then, the steady-state probability of having  $k$  type  $i$  orders is

$$\begin{aligned}
p^i(k) &= \sum_{\substack{y_1, \dots, y_m, j \neq i, y_i = k \\ 0 \leq y_j \leq N_j + S_j}} p(y_1, y_2, \dots, y_m), \tag{10}
\end{aligned}$$

which is used to compute the costs expressed in Eq.s (1-2). Note that the probabilities  $p^i(k)$  as well as the resulting costs would change with any change in any  $S_i$ ,  $i = 1, \dots, m$ . Therefore, the minimum system costs in Eq.s (1-2) have to be found by trying different  $(S_1, S_2, \dots, S_m)$  vectors.

### 3.2 The HF Policy: The Hybrid FCFS System

In this section, we study the system where a shared inventory of  $S > 0$  spare parts is kept for all fleets in addition to a reserved inventory of  $S_i \geq 0$  spare parts for each fleet  $i$ /type  $i$  machines,  $i = 1, \dots, m$ . When a type  $i$  machine fails, the failed component is sent to the repair shop. If there is positive stock in the shared inventory, one of the spare parts is installed on that machine, thus avoiding any machine stoppages. If the shared inventory happens to be empty but the reserved inventory level is positive, a spare part from the reserved inventory is used. Otherwise, since using a spare part from the reserved inventory of another fleet is

prohibited, the type  $i$  machine is down until a repaired component can be sent from the repair shop.

Let  $O(t)$  be the number of components in the repair shop at time  $t$ . If  $O(t) \leq S$ , the shared inventory level is  $I(t) = S - O(t)$  spare parts. All reserved inventories are at their respective base-stock levels  $S_i$ , and  $N_i$  machines are operational in each fleet. Therefore, whenever a component is repaired it is placed in the shared inventory, raising its level by 1. Letting  $O_i(t)$  denote the number of type  $i$  orders at time  $t$ , when  $O(t) \leq S$ ,  $O_i(t) = 0 \forall i$ .

We assume w.l.o.g that  $O(0) = 0$ . Letting  $\varsigma_0^D = 0$ , we define the following stopping times,

$$\begin{aligned}\varsigma_m^U &= \inf \{t : O(t) = S | t > \varsigma_{m-1}^D\}, \\ \varsigma_m^D &= \inf \{O(t) = S - 1 | t > \varsigma_m^U\}.\end{aligned}\tag{11}$$

In other words,  $\varsigma_m^U$  is a failure instant (equivalently, an arrival instant) of a component (sent to the repair shop) when the shared inventory level decreases from 1 to 0, and  $\varsigma_m^D$  is a repair completion instant when the shared inventory level increases from 0 to 1 for the  $m$ th time since time 0. Thus,  $D = \bigcup_{m=1}^{\infty} [\varsigma_m^U, \varsigma_m^D)$  is the time period during which each additional failed component sent by a fleet  $i$  would be a type  $i$  order for a repaired component, either to fix a down machine or to increase its reserved inventory level (which is below  $S_i$ ) by 1. Note that the system during  $D$  is probabilistically identical to the system studied in Section 3.1. While the system is in  $D$ , if  $O(t) = S$ , this corresponds to the idle period of a RIF system, with  $N_i$  machines operational, reserved inventory at  $S_i$ , and  $O_i(t) = 0, \forall i$ . At other times in  $D$  when  $O(t) > S$ ,  $O_i(t) > 0$  for some  $i$ . When a repair is done, the component is sent to the fleet with the longest standing order. As before, if  $O_i(t) \leq S_i$ , the reserved inventory level for type  $i$  machines is  $I_i(t) = S_i - O_i(t)$ , all  $N_i$  machines are operational, and if  $S_i < O_i(t) \leq N_i + S_i$ , fleet  $i$  has  $O_i(t) - S_i$  down machines, i.e., components backordered. If  $p_D$  is the proportion of time the HF system is in  $D$ , the steady-state probability of having  $k$  type  $i$  orders in the HF system will be  $p_D p^i(k)$  where  $p^i(k)$  is given by Eq. (10).

Letting  $p(k) := P(O = k)$  be the steady-state probability of having  $k$  components in the

repair shop, the objective function can be expressed as follows:

$$C_{HF}^* \equiv \min_{\mathbf{S}} \{C(\mathbf{S}) : C(\mathbf{S}) = C_0(S) + p_D \sum_{i=1}^m C_i(S_i)\}, \quad (12)$$

where  $\mathbf{S} = (S, S_1, S_2, \dots, S_m)$ ,  $C_i(S_i)$  is the cost due to fleet  $i$  given by Eq. (2) and

$$C_0(S) = h \sum_{k=0}^S (S - k)p(k), \quad (13)$$

which is the expected holding cost in the shared inventory. All that remains is to compute  $p_D$ . To obtain this probability, we will consider the system when the number of orders is less than or equal to  $S$ . The system behaves as a birth-and-death process with the following local balance equations

$$\begin{aligned} \Lambda p(k) &= \mu p(k+1), \quad k = 0, \dots, S-2 \\ \Lambda p(S-1) &= \mu p(S) = \mu p_D p_0(0), \end{aligned}$$

where, as before,  $\Lambda = \sum_{i=1}^m N_i \lambda_i$ , and  $p_0(0)$  is found from Eq. (4). After expressing all  $p(k)$  in terms of  $p_D p_0(0)$  as

$$p(k) = r^{S-k} p_D p_0(0), \quad k = 0, \dots, S, \quad (14)$$

where  $r = \mu/\Lambda$ , using  $\sum_{k=0}^{S-1} p(k) = 1 - p_D$ , we obtain

$$p_D = \frac{1}{1 + p_0(0) \sum_{k=1}^{S-1} r^k}. \quad (15)$$

Again the probabilities,  $p(k)$ ,  $p_D$  and  $p^i(k)$  as well as the resulting costs would change with any change in any  $S$  or  $S_i$ ,  $i = 1, \dots, m$ . Therefore, the minimum system costs in Eq.s (12-13) have to be determined by trying different  $(S, S_1, S_2, \dots, S_m)$  vectors. Obviously, if we set  $S = 0$ , the HF Policy becomes the RIF Policy.

### 3.3 The SIF Policy: The FCFS System with a Shared Inventory

In this section, we study the system where only a shared inventory of  $S > 0$  spare parts is kept. When a type  $i$  machine fails, the failed component is sent to the repair shop. If

there is positive stock in the shared inventory, one of the spare parts is installed on that machine avoiding any machine stoppages. Otherwise, the type  $i$  machine is down until a repaired component can be sent from the repair shop. The SIF Policy is a special case of the HF Policy if  $S_i = 0, \forall i$  in Section 3.2. Using the definitions from Section 3.2, if  $O(t) \leq S$ , the inventory level is  $I(t) = S - O(t)$ , and all  $N_i$  machines are operational in each fleet. Therefore, whenever a component is repaired, it is placed in the inventory, raising its level by 1. Similarly, when  $O(t) \leq S, O_i(t) = 0 \forall i$ . When  $O(t) > S, O_i(t) > 0$  for some  $i$ , and fleet  $i$  has  $O_i(t)$  down machines, i.e., components backordered. When a repair is completed, the component is sent to the fleet with the longest awaiting order. The steady-state probability of having  $k$  type  $i$  orders (or backordered components) in the SIF system will be  $p_D p^i(k)$  where  $p_D$  from Eq. (15) and  $p^i(k)$  from Eq. (10) are computed by setting  $S_i = 0, \forall i$ .

Our objective in this problem is to determine the optimal  $S^*$  to minimize the long-run average system cost as a time-average, which is

$$C_{SIF}^* \equiv \min_S \{C(S) : C(S) = h \sum_{k=0}^S (S - k)p(k) + p_D \sum_{i=1}^m b_i \sum_{k=1}^{N_i} k p^i(k)\}, \quad (16)$$

where, as before,  $p(k)$  is the steady-state probability of having  $k$  components in the repair shop, which is obtained from Eq. (14). Since the probabilities  $p(k), p_D$  as well as the resulting costs depend on  $S$ , the minimum system cost in Eq. (16) has to be found by trying different  $S$  values.

The following result is due to the fact that both the RIF and SIF policies are special cases of the HF policy.

**Result 1** *Among the three FCFS policies, we have*

$$C_{HF}^* \leq \min\{C_{SIF}^*, C_{RIF}^*\}.$$

Result 1 provides theoretical support for the use of the HF policy rather than the RIF and SIF policies.

## 4 Systems with Priority Dispatching Policies

In this section, we will analyze four systems with priority dispatching policy. We will start with the RIP policy in Section 4.1 since its results are necessary for analyzing the HP policy in Section 4.2. Finally, the SP policy and IR policy with three rationing levels, being special cases of the HP policy, will be studied in Sections 4.3 and 4.4, respectively. The model in Section 4.1 can be employed to obtain the distribution of the number of type  $i$  customers in single server Markovian queueing systems where priority classes have state-dependent arrival rates.

### 4.1 The RIP Policy: The Priority System with Reserved Inventories

The RIP system is very similar to the RIF system: A separate inventory with a base-stock level  $S_i \geq 0$  is reserved for each fleet  $i$ ,  $i = 1, \dots, m$ . As soon as a type  $i$  machine fails, the failed component and a type  $i$  order are sent to the repair shop. If there is stock in its inventory, a spare part is installed on the failed type  $i$  machine. Otherwise, this machine stays out of operation until a repaired component can be sent. Therefore, the inventory level or the backorders of fleet  $i$  can be inferred from the number of type  $i$  orders, as in the RIF system. Thus, Eq.s (1-2) can be used to obtain the optimal cost,  $C_{RIP}^*$ , of the RIP system. The basic difference between the two policies arises in  $p^i(k)$ 's due to different dispatching policies. While in the RIF system, the repaired component is sent to the fleet (or placed in its reserved inventory) with the longest standing order, in the RIP system it is sent to the highest-priority fleet (or placed in its reserved inventory) among those with outstanding orders.

Determining the priority ranking of fleets in this problem without computing the cost is not straightforward. Using our method, we must try different ways of prioritization and choose the one that minimizes the costs. To discuss how  $p^i(k)$ 's can be computed in the

RIP system, w.l.o.g, we will assume that fleets/classes 1 to  $m$  are prioritized from highest to lowest.

Bitran and Caldentey (2002) provide a method to obtain  $p^i(k)$ 's for a two-class preemptive-priority system with state-dependent Poisson arrival rates possibly with class specific exponential service times. In Appendix A, we adjust Bitran and Caldentey's model for our problem. For two fleets, for a given  $(S_1, S_2)$ , and using Eq.s (A.17-A.18) in Eq.s (1 and 2), the average cost can be computed.

We will now analyze the problem when there are more than two fleets. We will compute  $p^i(k)$ 's in a recursive manner by adding a new class one at a time. Each time a new class is added, we will use Bitran and Caldentey's model for two-priority classes. Recalling from Appendix A that  $M_i = N_i + S_i$ , the following Theorem states how  $p^m(k)$  is found, given  $p^i(k)$  for  $i = 1, \dots, m - 1$ :

**Theorem 1** *Given  $p^i(k)$  for  $i = 1, \dots, m - 1$ ,  $p^m(k)$  ( $m > 2$ ) is equal to  $p^2(k)$  in Eq. (A.18) of a two-class RIP system with  $N_1 = 1$ ,  $S_1 = \infty$ ,  $\lambda_1 = \sum_{i=1}^{m-1} \bar{\Lambda}_i$  where  $\bar{\Lambda}_i = \sum_{k=0}^{M_i} \Lambda_i(k) p^i(k)$  and  $N_2 = N_m$ ,  $S_2 = S_m$ ,  $\lambda_2 = \lambda_m$ .*

**Proof.** We obtain  $p^i(k)$ 's for  $i = 1, 2$  according to Bitran and Caldentey, as explained in Appendix A. Assume that  $p^i(k)$ 's for  $i = 3, \dots, m - 1$ , have been found using the method described in Theorem 1. At time  $t$ , given that there are  $k$  repair orders from class  $i$ , the probability of failure in the next  $\Delta t$  time units is  $\Lambda_i(k) \Delta t$ . If we remove the condition on the number of repair orders at time  $t$ ,  $\bar{\Lambda}_i \Delta t = \sum_{k=0}^{M_i} \Lambda_i(k) p^i(k) \Delta t$  is the probability of a failure in fleet  $i$  in the next  $\Delta t$  time units. Then,  $\bar{\Lambda}_i$  is the average failure rate from fleet  $i$ , which is also the effective arrival rate of components from fleet  $i$  to the repair shop. Whether or not the arrival processes of components from different fleets are independent of each other,  $\sum_{i=1}^{m-1} \bar{\Lambda}_i$  is the total failure rate of the classes  $1, \dots, m - 1$ , which from the point of view of fleet/class  $m$  is a single high-priority class. Additionally, class  $m$  perceives a constant failure rate,  $\sum_{i=1}^{m-1} \bar{\Lambda}_i$ , for the single high-priority class while itself experiencing a state-dependent failure rate (for systems with finite and infinite population interactions, see, e.g. Boxma,

1986 and Kaufman, 1984). Then, we can use an equivalent system, i.e., the RIP system with two priority classes such that  $N_1 = 1$ ,  $\lambda_1 = \sum_{i=1}^{m-1} \bar{\Lambda}_i$  and  $S_1 = \infty$  (or  $S_1 = M$  where  $M$  is a large integer to guarantee that there is always one machine functional in fleet 1, and the failure rate is always  $\lambda_1$ ) and  $N_2 = N_m$ ,  $S_2 = S_m$ ,  $\lambda_2 = \lambda_m$ . In this case,  $p^2(k)$  of this equivalent RIP system gives  $p^m(k)$ . ■

## 4.2 The HP Policy: The Hybrid Priority System

The HP system is similar to the HF system in that a shared inventory of  $S > 0$  spare parts is kept for all fleets in addition to a reserved inventory of  $S_i \geq 0$  spare parts for each fleet  $i$ /type  $i$  machines,  $i = 1, \dots, m$ . When a type  $i$  machine fails, the failed component is sent to the repair shop. If there is positive stock in the shared inventory, one of the spare parts is installed on that machine avoiding any machine stoppages. If the shared inventory happens to be empty but the reserved inventory level is positive, a spare part from the reserved inventory is used. Otherwise, since using a spare part from the reserved inventory of another fleet is prohibited, the type  $i$  machine is down until a repaired component can be sent from the repair shop (permitting this results in the SP system studied in Section 4.3).

If  $O(t) \leq S$ , the shared inventory level is  $I(t) = S - O(t)$  spare parts. All reserved inventories are at their respective base-stock levels  $S_i$ , and  $N_i$  machines are operational in each fleet. Therefore, whenever a component is repaired, it is placed in the shared inventory, raising its level by 1. When  $O(t) \leq S$ ,  $O_i(t) = 0 \forall i$ . When  $O(t) > S$ , there are no spare parts in the shared inventory,  $O_i(t) > 0$  for some  $i$ , and fleet  $i$  has  $O_i(t)$  down machines, i.e., components backordered. Eq. (12) can be used to obtain the optimal cost,  $C_{HP}^*$ , of the HP system. The basic difference between the two policies arises in  $p^i(k)$ 's due to different dispatching policies when  $O(t) > S$ : While in the HF system, the repaired component is sent to the fleet with the longest awaiting order, in the HP system it is sent to the highest-priority fleet among those with outstanding orders. We again assume that fleets/classes 1 to  $m$  are prioritized from highest to lowest.

To analyze this system, we make use of the stopping times given in Eq. (11). This time the system during  $D$  (with  $p_D$  probability) is probabilistically identical to the system studied in Section 4.1. Using these  $p^i(k)$ 's from Theorem 1 with  $p_0(0) = \prod_{i=1}^m p^i(0)$  in Eq.s (15 -14), we can compute  $p_D$  and  $p(k)$ , and  $C_{HP}^*$  can be computed from Eq. (12).

Again the probabilities,  $p(k)$ ,  $p_D$  and  $p^i(k)$  as well as the resulting costs would change with any change in any  $S$  or  $S_i$ ,  $i = 1, \dots, m$ . Therefore, the minimum system costs in Eq.s (12-13) have to be found by trying different  $(S, S_1, S_2, \dots, S_m)$  vectors. Obviously, if we set  $S = 0$ , the HP Policy becomes the RIP Policy. Therefore, we can state the following result, which provides theoretical support for the use of the HP policy instead of the RIP policy:

**Result 2** *Between the HP and RIP policies, we have*

$$C_{HP}^* \leq C_{RIP}^*.$$

### 4.3 The Strict Priority (SP) Policy

The SP system is a special case of the HP system and is similar to the SIF system in that when a type  $i$  machine fails if there is positive stock in the shared inventory (with base-stock level  $S \geq 0$ ), one of the spare parts is installed on that machine, thereby avoiding any machine stoppages. If  $O(t) \leq S$ , there are  $S - O(t)$  spare parts in the inventory and all  $N_i$  machines are operational in each fleet. Therefore, whenever a component is repaired it is placed in the inventory, raising its level by 1. When  $O(t) > S$ , there are no spare parts in the inventory,  $O_i(t) > 0$  for some  $i$ , and fleet  $i$  has  $O_i(t)$  down machines, i.e., components backordered. Thus, Eq. (16) can be used to obtain the optimal cost,  $C_{SP}^*$ , of the SP system. The difference between the two policies is again in  $p^i(k)$ 's due to different dispatching policies when  $O(t) \geq S$ : In the SIF system, the repaired component is sent to the fleet with the longest standing order, and in the SP system it is sent to the highest-priority fleet among those with outstanding orders.

The SP system can be easily analyzed, as in Section 4.2, by setting  $S_i = 0$  for  $i = 1, \dots, m$ . Then  $C(S)$  can be computed via Eq. (16) for a given  $S$ , and a search on different  $S$  values

yields  $C_{SP}^*$ .

Unlike the RIP policy in Section 4.1, for certain cases the question of how to prioritize different fleets is known under the SP policy. Prioritization becomes effective only when there is no spare part in stock. Iravani and Kolfal (2005) in their Theorem 1 show that a preemptive priority is the optimal policy if  $b_j/\lambda_j \geq b_i/\lambda_i$ . Then fleet  $j$ /class  $j$  has a priority over fleet  $i$ /class  $i$ . Therefore, if this condition holds for the SP policy, one should give priority to fleet  $j$  over fleet  $i$ .

#### 4.4 The Inventory Rationing (IR) Policy

In this section, we will study the IR policy with three rationing levels with  $R_3 = S_{IR} \geq R_2 \geq R_1 = 0$  for a shared inventory with a base-stock level  $S_{IR} \geq 0$ . We assume that fleets/classes 1 to  $m$  are prioritized from highest to lowest, but observing priorities depends on the inventory levels. When  $R_2 < I(t) \leq R_3$ , a spare part is immediately installed on any machine that fails, and the failed component joins the repair shop queue. When a component is repaired, it is placed in the inventory, raising its level by 1. When  $I(t) \leq R_2$ , a spare part is used (if there are any) only when type 1 machine fails. If other types of machines fail, they are down until inventory level reaches  $R_2$ . When  $0 \leq I(t) < R_2$ , a repaired component is placed in the inventory, raising its level by 1, even if there are down machines of types 2,  $\dots$ ,  $m$ . When  $I(t) = R_2$ , if there are down machines of types 2,  $\dots$ ,  $m$ , a repaired component is sent to the highest priority fleet with down machines. If there are no down machines when  $I(t) = R_2$ , the repaired component is placed in the inventory. When  $I(t) \leq 0$ , any type 1 machine failure also results in a down type 1 machine. In this case, each repaired component is used for a down machine of type 1 if there are any. Otherwise, it is placed in the inventory.

Observe that the IR policy with three rationing levels can be analyzed as a special case of the HP policy studied in Section 4.2. If we set  $S = R_3 - R_2 - 1$  and  $S_1 = R_2$  and  $S_i = 0$  for  $i = 2, \dots, m$  in the HP system, the resulting cost for given  $R_3$  and  $R_2$  can be obtained.

And a search on  $R_3$  and  $R_2$  gives the optimal cost of such an IR policy.

However, the IR policy with more than three rationing levels cannot be analyzed as an HP system. Consider the case with four rationing levels,  $R_4 = S_{IR} > R_3 > R_2 \geq R_1 = 0$ . This policy is interpreted as follows: When  $R_3 < I(t) \leq R_4$ , a spare part is immediately installed on any machine that fails. When  $R_2 < I(t) \leq R_3$ , a spare part is only used for type 1 or type 2 machines that fail, while the others are down upon failure. When  $0 \leq I(t) < R_2$ , a spare part is only used for type 1 machine, and type 2 machines are also down upon failure. Therefore, if we attempt to use the HP system, even if we can set  $S = R_4 - R_3 - 1$ , it is not possible to set  $S_1 > R_2$  and  $S_2$ .

In the following result,  $IR_3$  refers to the IR policy with three rationing levels and hold true because the SP policy is a special case of the  $IR_3$  policy with  $R_3 \geq R_2 = R_1 = 0$ , and the  $IR_3$  policy is a special case of the HP policy.

**Result 3** *Among the HP, SP and  $IR_3$  policies, we have*

$$C_{HP}^* \leq C_{IR_3}^* \leq C_{SP}^*.$$

Result 3 provides theoretical support for the use of the HP policy rather than the SP and  $IR_3$  Policies.

## 5 Numerical Experiment

In previous sections, we have analyzed several spare part inventory systems with different dispatching rules, showing that for the FCFS and priority dispatching policies, hybrid models (HF and HP policies, respectively) with a shared inventory, possibly with reserved inventories for each fleet, minimize the system costs. However, we did not address several important questions. (i) In all these policies, we assumed a centralized repair shop. Alternatively, we could consider a separate repair shop and a separate inventory for each fleet, which we will call the *Base Case* BC system. 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), what is the benefit of repair shop pooling in our problem? *(ii)* How do the HF and HP system perform with respect to one another? *(iii)* How do the special cases of the HF and HP systems perform among themselves and with respect to the HF and HP systems? *(iv)* What happens to the relative performances of the HF, HP and IR, namely, best policies when transportation times are non-negligible?

The answers to these questions are important for an outsourcing company for which each fleet represents a different customer. Instead of reserving its repair resources separately for each client, possible benefits from repair shop pooling could be important. This will be analyzed in Section 5.2 while addressing Question *(i)*.

An outsourcing company is likely to prioritize its customers according to customer-specific penalties included in the contracts. As long as the outsourcing company is willing to incur the penalty (backordering cost), it can defer maintenance service (i.e., dispatching spare parts) for less important customers if this yields a system wide cost decrease compared to FCFS policies. To this end, we will address Questions *(ii)* and *(iii)* in Section 5.3.

Finally, when transportation times are non-negligible, customers and the outsourcing companies might agree to have the reserved inventories at the customers' locations while keeping the shared inventory at the repair shop facility of the outsourcing company. This might lower the cost for the outsourcing company compared to pooling all inventories at the repair shop. As will be explained in Section 5.4, including transportation times in the analysis is quite difficult and even a simulation study could be quite expensive. Focusing on one example, we will address Question *(iv)* in this section.

Before presenting our extensive numerical study in detail, we will summarize our findings. In regards to Question *(i)*, the results in Section 5.2 clearly demonstrate the advantage of pooling the repair shop capacity. All policies studied in Sections 3 and 4 result in less cost than the BC system.

In regards to Questions *(ii)* and *(iii)*, the results in Section 5.3 demonstrate that the HP

policy is equivalent to the IR policy with three rationing levels and they are the champions of all policies in terms of minimizing the system costs. If hybrid models will not be used, when backordering costs are the same, policies with shared inventories (either the SIF or the SP) should be implemented. When backordering costs are different and the repair shop utilization increases, the RIP policy usually outperforms the SIF, SP and RIF policies.

In regard to Question (*iv*), the results in Section 5.3 demonstrate that when transportation times are non-negligible the HP policy is the best policy to consider. The system costs increase with longer transportation times, however the increase is much slower under the HF and HP policies than the increase under the IR policy. We also see that the transportation times do not affect the optimal IR threshold levels, whereas the advantage of the HF and HP policy is due to their ability to shift stock to the reserved inventories, which should be kept as local inventories at the locations of the fleets.

## 5.1 Basic Experimental Setup

In order to investigate the questions raised at the beginning of Section 5, we have designed a series of numerical experiments involving two fleets. The BC system can be analyzed using standard models (e.g., Gross and Harris, 1998, p. 82–83) separately for two fleets  $I$  and  $II$  (in priority based policies, fleet 1 will be the high-priority class and could be either fleet  $I$  or  $II$ ). For fleet  $i = I, II$  with  $N_i$  machines having  $\lambda_i$  failure rate and  $\mu_i$  repair rate, optimal  $S_i^*$  and optimal cost  $C_i(S_i^*)$  can be found. The optimal cost of the BC system is  $C_{BC}^* = C_I(S_I^*) + C_{II}(S_{II}^*)$ . In the alternative policies with a pooled repair shop, we set  $\mu = \mu_I + \mu_{II}$ .

We considered the following parameters:

- In all cases, the holding cost rate is set to 1.
- We considered the following down time cost rates:  $(b_I, b_{II}) \in \{(10, 10), (80, 80), (10, 100), (100, 10), (10, 800), (800, 10)\}$ .

- As an approximate measure of the repair shop utilization, we set  $u = \lambda_I N_I / \mu_I = \lambda_{II} N_{II} / \mu_{II} \in \{0.5, 0.7, 0.9\}$  corresponding to low, medium and high levels of utilization.
- The fleet sizes are chosen as follows: When  $N_I = 100$ ,  $N_{II}$  varies from 10 to 100 with increments of 10 (corresponding to cases with large fleet sizes), and when  $N_I = 10$ ,  $N_{II}$  varies from 1 to 10 with increments of 1 (corresponding to cases with small fleet sizes).
- In the BC system, we always set  $\mu_I = 1$  and  $\lambda_I = u/N_I$ . For the second fleet, for all problems, we first set  $\mu_{II} = 1$  and  $\lambda_{II} = u/N_{II}$  and then  $\lambda_{II} = \lambda_I$  and  $\mu_{II} = \lambda_I N_{II} / u$ .

Thus, due to 3 different utilizations, 6 different  $(b_I, b_{II})$  values, 2 different  $N_I$  and 10 different  $N_{II}$  values, we have  $3 \times 6 \times 2 \times 10 = 360$  problems. And for each one of these 360 problems, we have either  $\mu_{II} = 1$  or  $\mu_{II} = \lambda_I N_{II} / u$ . This gives a total of 720 cases. We compute the optimal cost under the BC system as well as the 7 policies analyzed in Sections 3 and 4.

## 5.2 Benefits of Repair Shop Pooling

To answer Question (i) in addressing the benefit of repair shop pooling, for each of the 720 examples, for the FCFS policies

$$\Delta_{BC}^{RIF} \equiv \frac{C_{BC}^* - C_{RIF}^*}{C_{BC}^*}, \Delta_{BC}^{HF} \equiv \frac{C_{BC}^* - C_{HF}^*}{C_{BC}^*}, \Delta_{BC}^{SIF} \equiv \frac{C_{BC}^* - C_{SIF}^*}{C_{BC}^*},$$

and for the priority policies

$$\Delta_{BC}^{RIP} \equiv \frac{C_{BC}^* - C_{RIP}^*}{C_{BC}^*}, \Delta_{BC}^{HP} \equiv \frac{C_{BC}^* - C_{HP}^*}{C_{BC}^*}, \Delta_{BC}^{SP} \equiv \frac{C_{BC}^* - C_{SP}^*}{C_{BC}^*}, \Delta_{BC}^{IR} \equiv \frac{C_{BC}^* - C_{IR}^*}{C_{BC}^*}.$$

These ratios measure the cost decrease due to repair shop pooling under the policies introduced in Sections 3 and 4 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 with a centralized repair shop.

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

	Min(%)	Average(%)	Median(%)	Max(%)
$\Delta_{BC}^{RIF}$	15	39	38	69
$\Delta_{BC}^{HF}$	19	46	46	72
$\Delta_{BC}^{SIF}$	7	37	37	56
$\Delta_{BC}^{RIP}$	17	43	40	73
$\Delta_{BC}^{HP}$	24	51	50	75
$\Delta_{BC}^{SP}$	9	42	43	59
$\Delta_{BC}^{IR}$	24	51	50	75

### 5.3 Relative Performance of Policies

Before answering Questions (ii) and (iii), we start by stating that the optimal HP policy and the optimal IR policy coincided in all 720 examples. In other words, the optimal base-stock level under the HP policy was always  $S^* = R_3^* - R_2^* - 1$ , where  $R_3^*$  and  $R_2^*$  are the optimal inventory rationing levels under the IR policy. Both the HP and IR policies prioritized the same class. Moreover, the HP policy reserved an inventory for the higher priority class ( $S_1^* \geq 0$ ) but not for both classes and  $S_1^* = R_2^*$ . Thus, the optimal costs turned out to be the same for both. This also explains why  $\Delta_{BC}^{IR}$  and  $\Delta_{BC}^{HP}$  were the same in Table 1. More importantly, their optimal cost was the minimum for all 720 cases. Therefore, for the FCFS policies

$$\Delta_{RIF}^{HP} \equiv \frac{C_{RIF}^* - C_{HP}^*}{C_{RIF}^*}, \Delta_{HF}^{HP} \equiv \frac{C_{HF}^* - C_{HP}^*}{C_{HF}^*}, \Delta_{SIF}^{HP} \equiv \frac{C_{SIF}^* - C_{HP}^*}{C_{SIF}^*},$$

and for the priority policies

$$\Delta_{RIP}^{HP} \equiv \frac{C_{RIP}^* - C_{HP}^*}{C_{RIP}^*}, \Delta_{SP}^{HP} \equiv \frac{C_{SP}^* - C_{HP}^*}{C_{SP}^*}.$$

These ratios measure the cost decrease due to using the HP (or equivalently, IR) policy instead of the other policies studied in Sections 3 and 4.

Table 2: The minimum, average, median and maximum values of cost reduction due to the HP/IR policies.

	Min(%)	Average(%)	Median(%)	Max(%)
$\Delta_{RIF}^{HP}$	6	20	20	35
$\Delta_{HF}^{HP}$	0	10	9	31
$\Delta_{SIF}^{HP}$	0	21	19	66
$\Delta_{RIP}^{HP}$	3	14	13	26
$\Delta_{SP}^{HP}$	0	14	10	62

In Table 2,  $\Delta_{HF}^{HP}$  compares the best FCFS policy, namely the HF policy, with the best priority-based policy, namely the HP policy. The HP policy results in, on average, 10% less cost than the HF policy and never incurs more cost. Therefore, we recommend the HP policy as the best maintenance practice.

If the hybrid models or the IR policy are not used, we would have four traditional policies, i.e., the SIF and SP policies with shared inventory and the RIF and RIP with reserved inventories for each fleet. It is not clear from Table 2 what the relative performances of these four policies are, and we summarize this as follows:

- The cost of the best priority based policy is less than or equal to that of the best FCFS based policy:

$$\min\{C_{SP}^*, C_{RIP}^*\} \leq \min\{C_{SIF}^*, C_{RIF}^*\}.$$

- Between the SP and SIF policy, we observe:

$$C_{SP}^* \leq C_{SIF}^*.$$

Although we have not established a formal proof, we can use the following intuition to explain this observation. Both policies are FCFS policies as long as there is positive stock in inventory. However, when there is a shortage, the SP policy uses the static

preemptive-resume priority policy, which due to the problem parameters considered turn out to be the optimal dispatching policy (Iravani and Kolfal, 2005).

- When downtime cost rates are the same for both fleets, the HF policy, in almost all cases, turns out to be the SIF policy (it does not reserve inventories for the fleets). With equal downtime costs, we see that the SIF and SP policies are better than the RIF and RIP policies.
- When  $(b_I, b_{II}) = (10, 100)$  or  $(100, 10)$ , at low utilization level, i.e., when  $u = 0.5$ , the SIF policy is better than the RIF policy.
- When  $(b_I, b_{II}) = (10, 100)$  or  $(100, 10)$ , at high utilization level, i.e., when  $u = 0.9$ , the RIP policy is better than the SP policy.
- Although there are exceptions, when the backordering costs are different, the RIP policy is preferable to the RIF, SIF and SP policies.

## 5.4 Relative Performance of Policies in the Presence of Transportation Times

In this paper, we assume that the transportation times of spares and repaired components to the fleets and failed components to the repair shop are negligible. However, in reality this might not be the case. Even if the transportation costs are negligible, the transportation times affect the holding and down time costs of the system. In such cases, pooling inventories at the locations of the repair shop could increase the costs. Instead, reserved inventories discussed in Sections 3 and 4 could be kept at the locations of the fleets instead of keeping them at the repair shop location. Yet, including transportation times in these analytical models is quite difficult. One can use simulation to find the optimal base-stock levels and the cost, however, when the fleet sizes are big and repair shop utilization is high searching over the vector of base-stock levels can be quite time consuming.

Given this, we chose the following example setting from Section 5.1 to investigate the impact of non-negligible transportation times:  $N_I = N_{II} = 10$ ,  $(b_I, b_{II}) = (100, 10)$ ,  $\lambda_I = \lambda_{II} = 0.09$  and  $\mu = 2$ . We assumed that the repair shop and the shared inventory are located at an equidistant point from both fleets. Among the 7 policies, we only considered the HF, HP and the IR policies. Since Sahba and Balçioğlu (2010) show via numerical examples that repair shop pooling is not beneficial when the mean transportation times exceed the mean repair time, we varied the deterministic transportation time as a percentage of the mean repair time.

In this setting, if the HF and HP policies reserve inventories for the fleets, they are kept at the locations of the fleets. Moreover, the policies slightly change. When a type  $i$  machine fails for  $i = 1, 2$ , if there is stock in its reserved inventory without any delay one of the spare parts is installed on that machine avoiding any machine stoppages. At the same time, if there is stock in the shared inventory a spare part is dispatched to this reserved inventory and arrives at the destination after a deterministic transportation time. By the time the spare part arrives at the fleet, if the reserved inventory is found to be depleted and there is at least one down machine, it is immediately installed on one of the down machines. It takes the same amount of time for the failed component to be transported to the repair shop and join the queue. This implies that at certain time intervals both the shared and the reserved inventories can be positive and below their respective base-stock levels (when transportation times are negligible if the shared inventory is not zero, the reserved inventories are always at their base-stock levels). If the shared inventory is non-zero when a component is repaired, it is placed in the shared inventory because even if the reserved inventories can be below their base-stock levels at that instant, the outstanding stock will be already on its way. Under both policies, the dispatching of the spare parts from the positive shared inventory is done on an FCFS basis. They differ when the shared inventory is zero. In this case, the HF policy sends the repaired component to the fleet that has the longest awaiting order. The HP policy, on the other hand, dispatches the repaired component to the high-priority fleet among those with pending orders. We assume that if a component is being transported to

a fleet, it cannot change its destination.

We simulated the HF, HP and IR policies with non-negligible transportation times to find their optimal inventory control parameters and costs using a commercial software package Arena Version 12. We determined 240 time units as the warm-up period and used the batch-means approach with 20 batches, with each batch of length 4,320 time units. The maximum 95% confidence interval half-widths around the average system costs was less than 0.5, thus we do not present them in Table 3.

In Table 3, we vary the ratio of transportation time between the repair shop and the location of a fleet as the percentage of the mean repair time, which we denote as TT/MRT (transportation time/mean repair time). In this table, the costs when transportation times are 0 are obtained from the analytical models of this paper. When transportation times are non-negligible, the most expensive policy turns out to be the IR policy. In all three policies, longer transportation times increase the systems costs more. However, the HF and HP policies appear to be minimally affected by the transportation times. When we compare the costs of 100% case to the 0% case, the system costs under the HF and HP policies increase by 4% and 9%, respectively. Under the IR policy, on the other hand, the system cost increases by 333% from 13.6 to 58.7. Among the three policies, the HP policy appears as the best policy to implement.

Table 3: Comparison of the HF, HP, and IR policies when transportation times are non-negligible.

TT/MRP	0%	10%	20%	40%	60%	80%	100%
The IR Policy	13.6	18.1	22.8	32.1	41.2	50.0	58.7
The HF Policy	16.0	15.9	16.2	16.4	16.5	16.6	16.7
The HP Policy	13.6	13.7	13.9	14.1	14.2	14.2	14.7

When we look at the optimal inventory control parameters, we observe the following: The parameters of the optimal IR policy are unaffected by the transportation time and they

are  $R_3^* = 10$  and  $R_2^* = 3$  for all cases presented in Table 3. This is not really surprising because keeping higher threshold levels of stock would not provide a remedy for delays due to transportation times. This observation is most likely applicable to any policy that pools all inventories at the location of the repair shop. Our analytical models could be useful in determining robust inventory control parameters, which in the absence of analytical models, can be simulated to estimate the system costs.

In Table 4, we present how the vector  $(S^*, S_1^*, S_2^*)$  changes under the HF and HP policies with increasing transportation times. What we observe is that the total number of spare parts stocked tends to increase with the transportation times. The advantage of these policies is obviously due to their ability to shift stocks to the reserved inventories, which help them reduce down time costs.

Table 4: The optimal inventory control parameters  $(S^*, S_1^*, S_2^*)$  under the HF and HP policies.

Transportation Time	0%	10%	20%	40%	60%	80%	100%
The HF Policy	(10,6,0)	(8,7,1)	(9,7,1)	(9,7,1)	(9,7,1)	(8,8,2)	(9,7,1)
The HP Policy	(10,3,0)	(10,3,1)	(9,3,2)	(9,3,2)	(9,3,2)	(7,4,3)	(7,5,3)

## 6 Conclusion and Future Research

In this paper, we have analyzed several repair shop/inventory systems for a spare part provisioning problem. Our analysis shows that when transportation times and costs are negligible, pooling repair shop capacity always decreases the system cost. Among the policies with a centralized repair shop, the HP and IR systems have proven to be superior. This is an important observation especially for outsourcing companies with the potential to prioritize their clients. These results lead to other research interests that we are planning to work on. In this paper, the IR system with three rationing levels was expressed as a special case of the HP system. With more than three rationing levels, we cannot make the same conversion.

Therefore, both systems should be compared in settings that will yield more than three threshold levels in the IR system. Furthermore, if the IR system sustains its performance, it would be worthwhile to examine if it is also the optimal policy in systems with a centralized spare part inventory. Our results provide good guidance when all the inventories are stocked at the location of the repair shop. However, when transportation times are non-negligible between fleets and the repair shop, the HF and HP policies can keep the inventories they reserve for fleets at the locations of the fleets. This gives them the ability to face minimal cost increases with higher transportation times. Our simulation example from which we make this observation shows that the IR policy starts performing very poorly with longer transportation times. Thus, we envision that not only in the spare part provisioning setting but also in the production/inventory systems, which assume constant demand rates, analytical models considering transportation times would become quite important.

## Acknowledgements

This work was supported in part by Natural Sciences and Engineering Research Council (NSERC) of Canada.

## References

- Abboud, E. N. and J. N. Daigle. 1997. "A Little's Result Approach to the Service Constrained Spares Provisioning Problem for Repairable Items," *Operations Research*, Vol. 45, No. 4, 577–583.
- Abouee-Mehrzi, H., B. Balcioğlu, and O. Baron. 2010. "Strategies for a single product multi-class  $M/G/1$  make-to-stock queue", *under review*.
- Benjaafar, S., L. W. Cooper, J. S. Kim. 2005. "On the Benefits of Pooling in Production-Inventory Systems", *Management Science*, Vol. 51, No. 4, 548–565.
- Bhat, U. N. and G. K. Miller. 2002. *Elements of applied stochastic processes*, Wiley, New

Jersey.

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

Boxma, O. J. 1986. "A queueing model of finite and infinite source interaction," *Operations Research Letters* Vol. 5, 245–254.

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

Dshalalow, J. 1991. "On single-server closed queues with priorities and state dependent parameters", *Queueing Systems*, Vol. 8, 237–53.

Eppen, G. D. 1979. "Effects of centralization on expected costs in a multi-location newsboy problem", *Management Science*, Vol. 25, 498-501.

Graves, S. C and J. Keilson. 1983. "System Balance for Extended Logistic Systems", *Operations Research*, Vol. 31, 234–252.

Gross, D., H. D. Kahn, and J. D. Marsh. 1977. "Queueing models for spares provisioning", *Naval Research Logistics Quarterly*, Vol. 24, 521–526.

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

Ha, A. 1997a. "Inventory Rationing Policy in a Make-to-Stock Production System with Several Demand Classes and Lost Sales", *Management Science* Vol. 43, 1093–1103.

Ha, A. 1997b. "Stock-Rationing Policy for a Make-to-Stock Production System with Two Priority Classes and Backordering", *Naval Research Logistics*, Vol. 44, 457–472.

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.

- Kranenburg, A. A. and G. J. van Houtum. 2008. “Service differentiation in spare parts inventory management”, *Journal of the Operational Research Society*, Vol. 59, 946–955.
- Miller, D. R. 1981. “Computation of Steady-State Probabilities for  $M/M/1$  Priority Queues”, *Operations Research*, Vol. 29, 945–958.
- Sahba, P and B. Balcioglu. 2010. “A Model for Optimizing Repairshop Capacity and Spare Parts Inventory”, *working paper, University of Toronto*.
- Sanajian, N., H. Abouee-Mehrizi, and B. Balcioglu. 2010. “Scheduling policies in the  $M/G/1$  make-to-stock queue”, *Journal of the Operational Research Society*, Vol. 61, 115–123.
- 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. “An Application of the Birth and Death Process to the Provision of Spare Machines”, *Operational Research Quarterly*, Vol. 5, 95–108.
- 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. 2001. “Assessing the Benefits of Different Stock-Allocation Policies for a Make-to-Stock Production System”, *Manufacturing & Service Operations Management*, Vol. 3, 105–121.
- de Véricourt, F., F. Karaesmen, and Y. Dallery. 2002. “Optimal Stock Allocation for a Capacitated Supply System”, *Management Science*, Vol. 48, 1486–1501.
- Yu, Y., S. Benjaafar, and Y. Gerchak. 2009. “Capacity Sharing and Cost Allocation among Independent Firms in the Presence of Congestion”, *under review*.

## Appendix A The DP System with Two Classes

We remind from Section 3.1 that  $p(y_1, y_2) = P(O_1 = y_1, O_2 = y_2)$  is the steady-state probability of having  $y_1$  orders from higher priority class 1 and  $y_2$  orders from class 2. Similarly,



the right eigenvector of  $\mathbf{B}_0$  associated to eigenvalue 0, Bitran and Caldentey define another sequence of vectors  $\mathbf{C}_0 = \tilde{\mathbf{P}}$  and  $\mathbf{C}_k = \Lambda_2(k-1)\mathbf{B}_k^{-1}\mathbf{C}_{k-1}$ ,  $k = 1, \dots, M_2$ .

Then,

$$p^2(k) = \frac{\mathbf{e}^T \mathbf{C}_k}{\sum_{j=0}^{M_2} \mathbf{e}^T \mathbf{C}_j} \text{ for } k = 0, \dots, M_2. \quad (\text{A.18})$$