Spare Parts Provisioning for Multiple k-out-of-n:G Systems

Pedram Sahba
University of Toronto, Department of Mechanical and Industrial Engineering
5 King's College Rd., Toronto, ON M5S 3G8, Canada,
pedram@mie.utoronto.ca
Barış Balcıoğlu
Sabancı University, Faculty of Engineering and Natural Sciences,
Orhanlı-Tuzla, 34956 Istanbul, Turkey,
balcioglu@sabanciuniv.edu
Dragan Banjevic
University of Toronto, Department of Mechanical and Industrial Engineering
5 King's College Rd., Toronto, ON M5S 3G8, Canada,
banjev@mie.utoronto.ca

Abstract

In this paper, we consider a repair shop that fixes failed components from different k-out-of-n:G systems. We assume that each system consists of the same type of component; to increase availability, a certain number of critical components are stocked as spare parts. We permit a shared inventory serving all systems and/or reserved inventories for each system; we call this a hybrid model. Additionally, we consider two alternative dispatching rules for the repaired component. The destination for a repaired component can be chosen either on a first-come-first-served basis or by following a static priority rule. Our analysis gives the steady-state system size distribution of the two alternative models at the repair shop. We conduct numerical examples minimizing the spare parts held while subjecting the availability of each system to exceed a targeted value. Our findings show that unless the availabilities of systems are close, the HP policy is better than the HF policy.

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

1 Introduction

Many complex and technologically advanced systems such as those in the electrical power industry (Levitin and Amari, 2010) and equipment such as radar or sonar systems used in mining (**de Smidt-Destombes et al., 2004**) are k-out-of-n:G systems comprised of identical components. A k-out-of-n:G system consists of n components each of which can fail from time to time. The system is deemed functional/available as long as a minimum of k components are functional. In this paper, we model a repair shop that fixes failed components from several such systems with spares kept to increase the availability of these systems. Our research was motivated by a British Columbia based mining company that uses thickeners in its processes. A thickener is a large tank with a slow turning rake used to settle and remove precipitated solids. In the acidic, neutral, and clarifying parts of the process, different k-out-of-n:G systems are utilized, but each uses the same type of rake drive. For settings such as these, the questions of interest are: (i) Should we reserve a separate spare parts inventory for each system, or should a single spare parts inventory be shared by all systems, or would a mixture of the two be more cost effective? (ii) And, when a repaired component from the repair shop is dispatched, should we choose the destination system/reserved inventory according to a first-come-first-served (FCFS) rule or could prioritization of systems make this decision less costly? We develop two alternatives involving a mixed inventory topology (a shared inventory together with reserved emergency inventories) and a repair shop under the FCFS and priority-based dispatching policies. We obtain the steady-state system size distributions at the repair shop; this allows us to compare performance and address the questions raised above.

Considering repairable components in an k-out-of-n:G system with spares is not new in the literature. Generally, however, the focus is on a single system; therefore, dispatching policies (except for 1-out-of-n:G systems) or different inventory structures do not arise from the problem context. Gupta and Sharma (1981) model such a system as a Markov chain by introducing operational, repair and installation as possible states for a component that is not stored in inventory. Fawzi and Hawkes (1991) revisit the same problem and assume that the single repair server gives installation of a spare part preemptive repeat priority over repair. In our study, we consider instantaneous installations; instead, a component is either operational, in inventory, or in the repair shop (either waiting to be fixed or being repaired). Assuming also a single server to model the repair shop, Frostig and Levikson (2002) allow the repair times to follow non-Exponential distributions. In all these studies including ours, each broken component is sent to the repair shop as soon as it fails. **de Smidt-Destombes et al. (2004)**, on the other hand, assume that the repair process starts only after a given number of failed components accumulate.

The case of a 1-out-n:G system can represent a fleet of machines and has been extensively studied in the literature in machine interference or machine repairperson problems. However, including a joint repair shop is difficult even when an FCFS dispatching rule is followed for multiple 1-out-n:G systems. Earlier work 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. 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 as in Jaiswal (1968, p. 71,79). 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 et al. (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, a simple static non-preemptive priority policy is optimal, and they present 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.

In this paper, we analyze a Markovian single server queueing system with multiple classes of customers whose arrival rates are state-dependent. Under the FCFS and the priority 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. Then, we extend the model in Bitran and Caldentey (2002) 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 allow a flexible use of spare part inventory structures. This is important because earlier studies (Graves and Keilson, 1983, Dshalalow, 1991, Abboud and Daigle, 1997) with spare part inventories usually assume a single finite population. Therefore, reserved inventories for each population have not been compared to a shared inventory for all. **Benjaafar et al. (2005)** make this type of comparison in systems operating under the FCFS policy with constant customer arrival rates; they prove that a shared inventory results in a cost less than or equal to that of the alternative with reserved inventories when the holding cost is the same in both models and the backordering cost rate is the same for each class. In our case, different inventory levels change the state-dependent arrival rates, rendering comparison more difficult.

We make this comparison by developing two models serving multiple k-out-of-n:G systems. First, the HP (hybrid priority) model has a shared inventory for all systems and may have reserved inventories for some systems. The shared inventory is depleted on an FCFS basis, and when it is empty, the repair shop dispatches repaired components to systems/reserved inventories according to their priorities. Second, the HF (hybrid FCFS) model is similar except that when its shared inventory is depleted, the repair shop dispatches repaired components to the system (or the inventory reserved for that system) with the longest waiting repair order. Incorporating dynamic priority rules is overly complex, and it is not studied in this paper.

Since it is not possible to show theoretically when one policy is better than the other one, we provide the results of our extensive numerical study in Section 3. Our results indicate that the HP policy is better in most of the examples; however, lower repair capacity degrades its performance, and if the minimum availabilities set for systems are close, then, the HF policy is better.

The rest of the paper is organized as follows. In Section 2, we present our problem and the models we consider. In Section 2.1, we analyze the HF model. In Section 2.2, the HP model, in which different fleets are given different priorities, is modeled. In Section 3, we present numerical results comparing their relative performances of these policies.

2 Alternative Policies

We consider a centralized repair shop that serves m systems parameterized by $i = 1, ..., m$. Each system i is a k_i-out-of-n_i:G system comprised of identical components and is available if k_i or more components out of n_i are functional. Although components fail from time to time, they are repairable. Additionally, spare components are kept to increase the proportion of times these systems are available. Times to failure, that is the periods between installation of a new or repaired component in system i and the next failure instant of this component, follow an exponential distribution with rate λ_i (implying that each repair makes the component as good as new, and the failure rate only depends on the system using it). **Different failure rates can be due to the type of service a system renders or specific operating conditions they are subject to**. **(Note that if the times to failure distribution has an increasing failure rate, our model operating with exponential times to failure assumption may provide inaccurate approximations.)** 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 μ. If there is a stock of critical components kept as spare parts, a spare component can be installed immediately to replace the failed component. Otherwise, the number of functional components in that system decreases by 1. When only k_i components are **functional and there is** no spare available, and if one more component fails, system i fails and is down until a repaired component can be dispatched from the repair shop (during such down times, the remaining $k_i - 1$ components do not fail).

In other words, keeping spare part inventories might help increase system availability at the expense of incurring inventory holding cost. Separate inventories for each system can be reserved; or due to the same component used, a shared inventory can serve all systems. In this paper, we model a mixture of the two, with both shared and reserved inventories calling it the *hybrid* model. In the hybrid model, by setting shared or reserved inventory levels to zero, one can create a system of only reserved inventories or of only a shared inventory, respectively. Therefore, the optimal cost of the hybrid model cannot be strictly higher than the optimal cost of having solely reserved inventories or only a shared inventory.

The problem also involves a component allocation problem. In this paper, we study the FCFS and the priority policies which result in two alternative policies, the *hybrid FCFS* (HF) policy and the *hybrid priority* (HP) policy. In both cases, in addition to a reserved inventory for each system, there is a shared inventory for all systems, and each inventory operates according to a base-stock policy. First spare parts from the shared inventory are expended, and only when they are depleted, are the reserved inventories used. **Now consider the periods when the shared inventory is empty, and some reserved inventories are below their base-stock levels, or some systems do not have all of their components functional**. When this is the case, the repair shop has pending repair orders from systems missing functional components or spares in their reserved inventories. **Therefore, each time a repair is finished, a policy needs to be followed to dispatch the fixed component.** We study the HF policy in Section 2.1. In this policy, the repair shop dispatches the repaired components in an FCFS manner among systems with pending orders. Under the HP policy, studied in Section 2.2, the repaired component is used to serve the highest priority system among those with pending orders.

2.1 The Hybrid FCFS (HF) Policy

In this section, we analyze the model in which a shared inventory of $S > 0$ spare parts is kept for all systems in addition to (emergency) reserved inventory of $S_i \geq 0$ spare parts for each system i, $i = 1, ..., m$. When a component fails, it is sent to the repair shop. If there is positive stock in the shared inventory, a spare part is installed. If the shared inventory happens to be empty but the reserved inventory level is positive, a spare part from the latter is used. Otherwise, system i lacks one more component until a repaired one can be sent from the repair shop. When $k_i - 1$ functional components remain, system i fails and no more component failures can be observed until a repaired component can be sent from the repair shop on an FCFS basis.

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 components are functional in each system. Therefore, whenever a component is repaired, it is placed in the shared inventory, raising its level by 1.

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

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

In other words, ς_m^U is a failure instant (equivalently, an arrival instant) of a component (at the repair shop) when the shared inventory level decreases from 1 to 0, and ς_m^D is a repair completion instant when the shared inventory level increases from 0 to 1 for the mth time since time 0. Thus, $D = \bigcup_{m=1}^{\infty} [\varsigma_m^U, \varsigma_m^D)$ is the time period during which each additional component to fail in system i generates a type i repair order. Let $O_i(t)$ be the number of type i orders at time t. If $O_i(t) \leq S_i$, the reserved inventory level for system i is $I_i(t) = S_i - O_i(t)$, all n_i components are operational, and if $S_i < O_i(t) \leq n_i + S_i - k_i + 1$, system i lacks $O_i(t) - S_i$ components. When a repair is done, the component is sent to the fleet with the longest standing order. Let $p(k) := P(O = k)$ ($p^i_{HF}(k) := P(O_i = k)$) be the steady-state probability of having k components (k type i orders) in the repair shop, and p_D be the proportion of time the HF model is in D.

Consider a separate model, the *reserved inventory-FCFS* (RIF) model, with exactly the same parameters (e.g., systems served, failure and repair rates, base-stock levels of the reserved inventories) but no shared inventory ($S = 0$). Obviously the HF model during D is probabilistically identical to the RIF model. If we can obtain the steady-state probability of having k type i orders in the repair shop in the RIF model, denoted by $p^i(k)$, then, the steady-state probability of having k type i orders in the HF model is simply

$$p^i_{HF}(k) = p_D p^i(k). \tag{2}$$

Therefore, the analysis of the RIF model is necessary for the HF model. In the next section, we derive the steady-state distribution of the number of orders of each type in the repair shop of the RIF model. Then, we obtain p_D (and $p(k)$) in the HF model in Section 2.1.2. With them, Eq. (2) gives $p^i_{HF}(k)$, which together with $p(k)$, provides the steady-state distribution of the number of repair orders in the HF system.

2.1.1 Obtaining $p^i(k)$ in the RIF model

In order to obtain $p^i(k)$, we start by characterizing the state of the RIF model at an arbitrary time by the vector $(y_1, y_2, \ldots, y_m)$ with y_i being the number of type i orders present at the repair shop. Observe that the failure rate from system i depends only on the number of components functional in system i. We define this failure rate $\Lambda_i(y_i)$ as

$$\Lambda_i(y_i) = \begin{cases} n_i\lambda_i, & \text{if } 0 \le y_i \le S_i, \\ (n_i + S_i - y_i)\lambda_i, & \text{if } S_i \le y_i < n_i + S_i - k_i + 1, \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

Adding the failure rates from all systems, $\Lambda(y_1, y_2, \ldots, y_m) = \sum_{i=1}^{m} \Lambda_i(y_i)$, gives us the state-dependent arrival rate of repair orders at the single server queue when the system is in state $(y_1, y_2, \ldots, y_m)$.

Let us consider a state $\omega_{\mathbf{n}}^{\mathbf{j}}$ with n repair orders at the repair shop, i.e., $y_1+y_2+\cdots+y_m = n$. We use a superscript j because there are K such states differing from one another due to the ordering of repair orders, and the steady-state probability of being in state $\omega_{\mathbf{n}}^{\mathbf{j}}$, $p_n(\omega_{\mathbf{n}}^{\mathbf{j}}) = q(y_1, \ldots, y_m)$, is the same for all j, $j = 1, \ldots, K$, where

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

For each one of these K states, summing up local balance equations,

$$q(y_1, \ldots, y_i, \ldots, y_m)\Lambda_i(y_i) = \mu q(y_1, \ldots, y_i + 1, \ldots, y_m),$$

over all i, we obtain

$$\begin{aligned} q(y_1, \ldots, y_m) &= \frac{\mu}{\Lambda(y_1, \ldots, y_m)}(q(y_1 + 1, y_2, \ldots, y_m) + q(y_1, y_2 + 1, \ldots, y_m) \\ &\quad + \cdots + q(y_1, y_2, \ldots, y_m + 1)). \end{aligned} \tag{4}$$

However, we are interested in the probability of being in any one of these K states with y_1 orders of type 1, y_2 orders of type 2, ... , and y_m orders of type m. We denote this probability by $p(y_1, y_2, \ldots, y_m)$:

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

If we multiply both sides of Eq. (4) by K, we arrive at

$$\begin{aligned} p(y_1, y_2, \ldots, y_m) &= \frac{\mu}{(n+1)\Lambda(y_1, \ldots, y_m)}\{(y_1 + 1)p(y_1 + 1, y_2, \ldots, y_m) \\ &\quad +(y_2 + 1)p(y_1, y_2 + 1, \ldots, y_m) + \ldots \\ &\quad +(y_m + 1)p(y_1, y_2, \ldots, y_m + 1)\}. \end{aligned} \tag{5}$$

Using Eq. (5), we can express all $p(y_1, y_2, \ldots, y_m)$ in terms of p_N, which is the probability of having $N = \sum_{i=1}^{m}(n_i + S_i - k_i + 1)$ repair orders in the system (the maximum number of components that can be in the repair shop). After employing the normalization constraint,

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

we obtain $p_{_N}$ and all $p(y_1, y_2, \ldots, y_m)$ where $y_i \in \{0, 1, \ldots, n_i + S_i - k_i + 1\}$ for $i = 1, \ldots, m$.

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

$$p^i(k) = \sum_{\substack{y_1, \ldots, y_m,\ j \neq i,\ y_i = k \\ 0 \leq y_j \leq n_j + S_j - k_j + 1}} p(y_1, y_2, \ldots, y_m). \tag{6}$$

2.1.2 Obtaining p_D in the HF model

To obtain p_D, we 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

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

where, $\Lambda = \sum_{i=1}^{m} n_i \lambda_i$, and $p_0(0)$ denotes the probability that the repair shop is idle and n_i components are running and S_i spare parts are available in the inventory for each class i in the RIF system. With $p^i(0)$'s obtained from Eq. (6), the probability $p_0(0)$ is found as

$$p_0(0) = \prod_{i=1}^{m} p^i(0).$$

After expressing all $p(k)$ in terms of $p_D p_0(0)$ as

$$p(k) = r^{S-k} p_D p_0(0), \quad k = 0, \ldots, S, \tag{7}$$

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

$$p_D = \frac{1}{1 + p_0(0) \sum_{k=1}^{S-1} r^k}. \tag{8}$$

With p_D in Eq. (8) and $p^i(k)$ in Eq. (6), we can compute $p^i_{HF}(k)$ in Eq. (2). Note that all these probabilities would change with any change in any S or S_i, $i = 1, \ldots, m$.

2.2 The Hybrid Priority (HP) Model

The HP model is similar to the HF model except for the dispatching policy employed during the period D (defined by making use of the stopping times given in Eq. (1)) when $O(t) > S$, and $O_i(t) > 0$ for some i. While in the HF model, the repaired component is sent to the system with the longest awaiting order, in the HP model, it is sent to the highest-priority system among those with **pending** orders. We assume that systems/classes 1 to m are prioritized from highest to lowest.

To analyze this system, we consider a separate model, called the *reserved inventory-priority* (RIP) model, with exactly the same parameters as the HP model but with no shared inventory ($S = 0$). The HP model during D is probabilistically identical to the RIP model. We obtain $p^i(k)$'s of the RIP model in Section 2.2.1, and with $p_0(0) = \prod_{i=1}^m p^i(0)$ in Eqs. (8 –7), we find p_D and $p(k)$ in the HP model. By substituting these in Eq. (2), we obtain the steady-state probability of having k type i orders in the HP model, $p^i_{HP}(k)$.

2.2.1 Obtaining $p^i(k)$ in the RIP model

We use a matrix approach similar to Bitran and Caldentey (2002) who 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. First, we adjust Bitran and Caldentey's solution to our problem.

As in Section 2.1.1, let $\Lambda_i(y_i)$ be the failure rate for class i for $i = 1, 2$ given that there are y_i orders.

Let $M_i = n_i + S_i - k_i + 1$, for $i = 1, 2$. Since for class 1 customers,

$$\Lambda_1(k)p^1(k) = \mu p^1(k+1), \text{ for } k = 0, \cdots, M_1 - 1,$$

holds, the sequence $\pi_0 = 1$ and $\pi_k = (\Lambda_1(k-1)/\mu)\pi_{k-1}$ can be defined such that

$$p^1(k) = \frac{\pi_k}{\sum_{j=0}^{M_1} \pi_j}, \text{ for } k = 0, \cdots, M_1. \tag{9}$$

For class 2, their algorithm is more complex: For a given k for the number of class 2 orders, we define

$$
\mathbf{A_k} = \begin{bmatrix}
a_{0,k} & -\mu & & & & 0 & \\
-\Lambda_1(0) & a_{1,k} & -\mu & & & & \\
& -\Lambda_1(1) & a_{2,k} & -\mu & & & \\
& & \cdots & \cdots & \cdots & & \\
& & & & -\Lambda_1(M_1-2) & a_{M_1-1,k} & -\mu \\
& & 0 & & & -\Lambda_1(M_1-1) & a_{M_1,k}
\end{bmatrix},
$$

where $a_{y_1,k}$ is given by

$$
a_{y_1,k} = \begin{cases}
N_1\lambda_1 + N_2\lambda_2, & y_1 = k = 0, \\
N_1\lambda_1 + \Lambda_2(k) + \mu, & y_1 = 0,\ k > 0, \\
\Lambda_2(k) + \mu, & y_1 = M_1, k \geq 0, \\
\Lambda_1(y_1) + \Lambda_2(k) + \mu & \text{otherwise.}
\end{cases}
$$

Additionally $(M_1+1) \times (M_1+1)$ matrices $\mathbf{B_k} = \mathbf{A_k} - \Lambda_2(k)(\mathbf{e_1 e^T})$ are defined, where $\mathbf{e_1}$ is an $(M_1+1) \times 1$ vector with 1 as its first entry and 0's for the rest, and $\mathbf{e^T}$ is the transpose of an $(M_1+1) \times 1$ vector of 1's. Except for $\mathbf{B_0}$, which has a rank M_1, all matrices $\mathbf{B_k}$ have full rank. Using $\tilde{\mathbf{P}}$, which is 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 $\mathbf{C_k} = \Lambda_2(k-1)\mathbf{B_k}^{-1}\mathbf{C_{k-1}}$, $k = 1, \cdots, M_2$.

Then,

$$
p^2(k) \;=\; \frac{\mathbf{e^T C_k}}{\sum_{j=0}^{M_2} \mathbf{e^T C_j}} \text{ for } k = 0, \cdots, M_2. \tag{10}
$$

We now extend the result to m classes. We compute $p^i(k)$'s in a recursive manner by adding one new class at a time. Each time a new class is added, we use the two-priority class model. With $M_i = n_i + S_i - k_i + 1$, the following Theorem states how $p^m(k)$ can be found, given $p^i(k)$ for $i = 1, \ldots, m-1$:

Theorem 1 *Given $p^i(k)$ for $i = 1, \ldots, m-1$, $p^m(k)$ $(m > 2)$ is equal to $p^2(k)$ in Eq. (10) of a two-class RIP system with $n_1 = k_1 = 1$, $S_1 = \infty$, $\lambda_1 = \sum_{i=1}^{m-1} \bar{\Lambda}_i$ where $\bar{\Lambda}_i = \sum_{k=0}^{M_i} \Lambda_i(k) p^i(k)$ and $n_2 = n_m$, $k_2 = k_m$, $S_2 = S_m$, $\lambda_2 = \lambda_m$.*

Proof. We obtain $p^i(k)$'s for $i = 1, 2$ according to two-class priority model. Assume that $p^i(k)$'s for $i = 3, \ldots, m-1$, 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 Δt 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 system i in the next Δt time units. Then, $\bar{\Lambda}_i$ is the mean failure rate from system i; it is also the effective arrival rate of components from system i at 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 $1, \ldots, m-1$, which from the point of view of system/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 it is 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 $n_1 = k_1 = 1$, $\lambda_1 = \sum_{i=1}^{m-1} \bar{\Lambda}_i$ and $S_1 = \infty$ (or $S_1 = M$ where M is a large integer to guarantee that there is always one component functional in system 1, and the failure rate is always λ_1), and $n_2 = n_m$, $k_2 = k_m$, $S_2 = S_m$, $\lambda_2 = \lambda_m$. In this case, $p^2(k)$ of this equivalent RIP system gives $p^m(k)$. ■

3 Numerical Experiment

In previous sections, we have analyzed the HF and HP policies. However, we have not compared the performances of the HF and HP systems. In order to investigate whether we can have general insight into when to use one policy instead of the other, we have designed a series of numerical experiments involving two k_i-out-of-n_i:G systems. These experiments are set up as optimization problems in which the minimization of the total capital cost tied up in keeping spares is the objective function (e.g., Louit et al., 2011, and Sahba and Balcıoğlu, 2011) and the steady-state availability of each fleet i has to meet a minimum target level A_i. Let $A_{P,i}(S, S_I, S_{II})$ denote the steady-state availability of system i under policy P which is

$$A_{P,i}(S, S_I, S_{II}) = 1 - p^i_P(n_i + S_i - k_i + 1), \;\; i = I, II.$$

Note that under the HP policy, the high-priority system/class 1 could be either system I or II. Also observe that $p^i_P(n_i + S_i - k_i + 1)$ is a function of S, S_I, and S_{II}, as well as the policy, and is found from Eq. (2) for the HF policy (and for the HP policy, with adjustments as explained at the beginning of Section 2.2). The optimization problem for HF and HP policies becomes

$$\min \;\; hS + h_I S_I + h_{II} S_{II},$$

subject to

$$A_{P,i}(S, S_I, S_{II}) \geq A_i, \;\;\; i = I, II,$$

where h, h_I, h_{II} are the holding cost rates due to the capital cost tied up in keeping spares in shared and reserved inventories, respectively.

Modeling optimization problems of this type is common in the literature (**Sasaki et al., 1977, Yanagi et al., 1981**). A common technique to solve these models is presented by Lawler and Bell (1966). Two other methods are dynamic programming (e.g., Messinger and Shooman, 1970), and **iteratively adding a spare to the system that results in the highest improvement in the system reliability per dollar spent (e.g., Barlow and Proschan, 1965, p. 162)**. One necessary condition to apply these methods is that $A_{P,i}(S, S_I, S_{II})$ must be monotone non-decreasing in each of the variables S, S_I, and S_{II}. However, while $A_{P,i}(S, S_I, S_{II})$ is non-decreasing in S_i, it can be shown to be non-increasing in S_j ($j \neq i$): Adding a unit of spare to the reserved inventory of system i may only improve its availability, and the proportion of time that system i is down may shorten. Hence, **the failure rate** of this system increases, resulting in higher utilization of the repair shop, which, in turn, lowers the availability of the other fleet. This fact can be used to skip some sub-optimal solutions in an exhaustive search to find the optimal solution. For instance, let S^L_{II}

and S_{II}^{L+} denote the lowest values of S_{II} that satisfy the availability constraints together with the set $\{S, S_I\}$ and $\{S, S_I+1\}$, respectively. Then, $S_{II}^{L+} \geq S_{II}^{L}$, and therefore, $\{S, S_I+1, S_{II}^{L+}\}$ cannot be the optimal solution. Note that this approach can be easily extended to problems involving more than two k-out-of-n:G systems.

3.1 The Summary of the Numerical Results

This section provides a brief summary of the numerical examples conducted to explore how the HF and HP policies perform with respect to one another. In all numerical examples discussed in this and subsequent sections, system I is a 90-out-of-100:G system ($k_I = 90$ and $n_I = 100$) with $A_I = 0.999$, $\lambda_I = 0.009$. The holding cost rates of all inventories are set to 1. By varying a certain parameter of system II or the repair rate, we have generated four sets of examples, each consisting of 600 examples, to be discussed in more detail in Section 3.2. 300 out of 2,400 were repeating examples, therefore, we ended up with a total of 2,100 examples.

In order to measure the cost decrease due to using the optimal HP policy instead of the optimal HF policy for 2,000 examples (100 out of 2,100 had an optimal cost of 0 under the HF policy), we computed

$$\Delta_{HF}^{HP} \equiv \frac{C_{HF}^{*} - C_{HP}^{*}}{C_{HF}^{*}}.$$

Table 1: The minimum, mean, median and maximum values of cost reduction due to the HP policy.

	Min(%)	Mean(%)	Median(%)	Max(%)
Δ_{HF}^{HP}	-800	38	67	100

From Table 1, we see that although the HP policy results in, on average, 38% less cost than the HF policy, in some cases it can be costlier (up to 800% of the cost of the HF policy). Therefore, in the next section, we compare the two policies in more detail.

3.2 The Relative Performance of the HF and HP Policies

As stated at the beginning of Section 3.1, we generate four sets of examples by choosing a certain parameter of system II or the repair rate, and assigning the parameter chosen 6 different values. For each parameter value, we increment the target availability for system II in steps of 0.001 such that $A_{II} \in \{0.9, 0.999\}$. In other words, in each set, for each parameter value considered, we obtain the optimal solution under the HF and HP policies for 100 different A_{II} values.

3.2.1 The Impact of Repair Capacity

In the first set of 600 examples, system II is also a 90-out-of-100:G system ($k_{II} = 90$ and $n_{II} = 100$) with $\lambda_{II} = 0.009$, i.e., identical to system I. With $u \in \{0.75, 0.8, 0.85, 0.9, 0.95, 0.99\}$, we vary the repair shop capacity μ as $\mu = 1.8/u$.

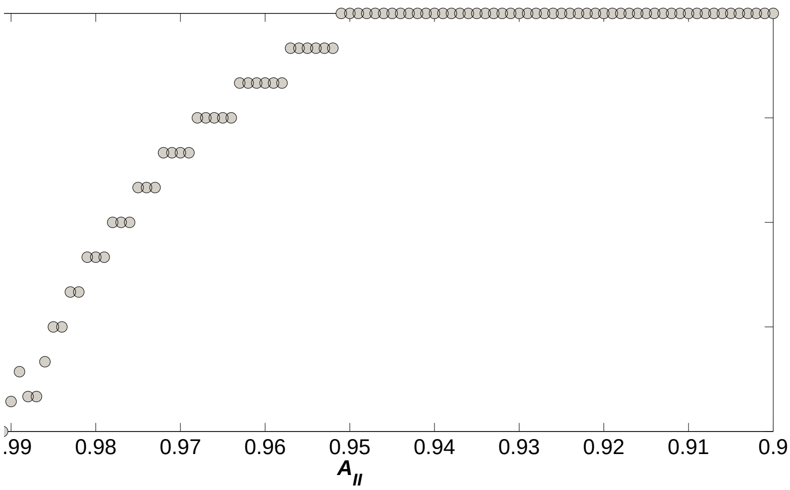

Figure 1: The cost reduction of the HP system ($\Delta_{HF}^{HP}\%$) compared to the HF system when $u = 0.9$

In Figure 1, $u = 0.9$ and $\mu = 2$. We see that the HF policy outperforms the HP policy for $A_{II} \geq 0.978$ only (negative values indicate that the HF policy outperforms the HP policy). Thus, the HF policy should be preferred only when target availabilities of identical systems are close. For the examples presented in Figure 1, the HP policy gives priority to system I and stores spare parts – if any – solely in the reserved inventory of system II. For $A_{II} \leq 0.951$, the HP policy does not carry any inventory at all. The HF policy, in contrast, stores spares in the shared and system I reserved inventories, and no matter how low A_{II} gets, the inventory levels do not reduce to 0 (hence, $\Delta_{HF}^{HP}\%$=100 for $A_{II} \leq 0.951$).

Let $\overline{A}_i$ ($\geq A_i$) denote the actual availability system i is provided with when the optimal number of spares is obtained under a given policy. In Figure 2, we see that under both policies $\overline{A}_{II}$ tends to decrease with A_{II} getting smaller until a minimum is met. This minimum is 0.951 for the HP policy, the availability at which it also starts carrying no inventory. The minimum actual availability under the HF policy, on the other hand, does not decrease below 0.972. At first glance, having a higher actual availability seems better, but from Figure 1, we recall that the HF policy incurs non-zero cost of carrying spares inventories. In other words, higher actual availability under the HF policy comes with a cost.

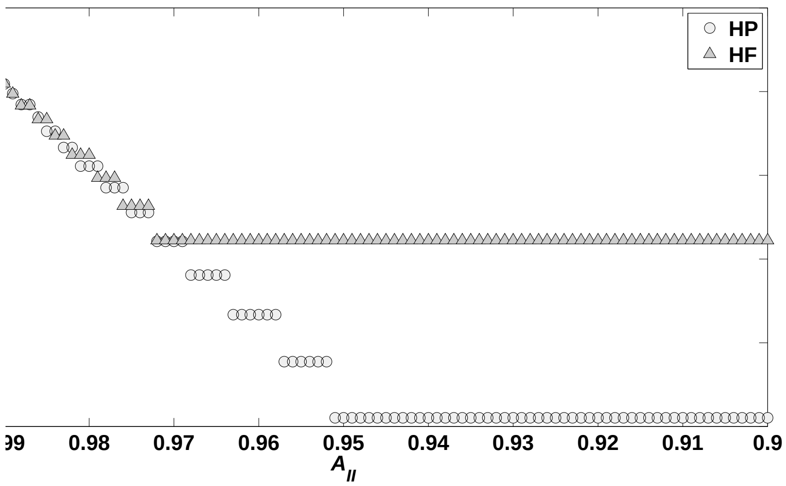

Figure 2: The actual $\overline{A}_{II}$ vs. target A_{II} availabilities for system II when $u = 0.9$

For other u or equivalently μ values, the relative performances of the two policies, as well as the way inventories are used, remain the same: system I is the high-priority class under the HP policy that stores all spares – if any – in system II reserved inventory. In contrast, the HF policy stores the spares in the shared inventory when the target availabilities of the systems are close. When A_{II} diminishes, the shared inventory level decreases while the reserved inventory of system I increases, but their sum, i.e., the total number of spares, reduces.

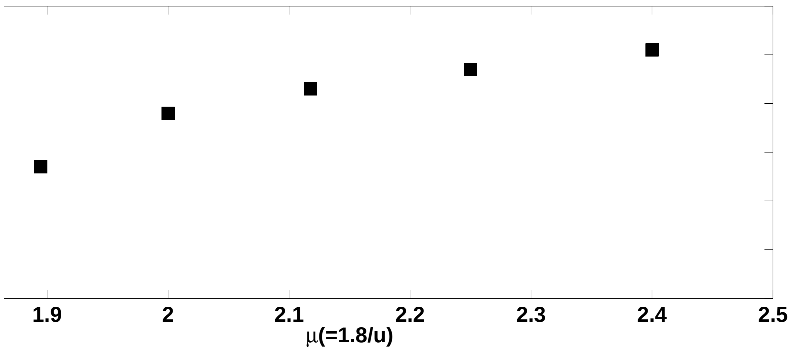

Figure 3: The maximum A_{II} value below which the HP policy outperforms the HF policy vs. μ_{II}

With lower μ, not only do the optimal levels of spares increase, but also the performance of the HP policy worsens. To see this, let T denote the A_{II} at which the performances (the optimal costs) of the HP and HF policies are the same. This means that when $A_{II} < T$, the HP policy is more cost-effective. Figure 3 plots the T values for six μ values. At the smallest capacity considered, when $\mu_{II} = 1.818$, we read $T = 0.949$. This implies that from the 100 examples optimized for both policies, the HP policy was more cost-effective than the HF policy in 49 examples when $A_{II} \in \{0.9, 0.948\}$. Figure 3 shows that with higher μ, T increases monotonically. At the highest capacity considered, when $\mu_{II} = 2.4$, the HP policy is better in 91 examples out of 100 for $A_{II} < 0.991 = T$. We conclude that if the repair capacity is low, A_{II} should be considerably smaller than $T < A_I = 0.999$ in order for the HP policy to beat the HF policy. With sufficiently high capacity, the HP policy performs better than the HF policy, even when the difference between A_I and A_{II} is not significant.

3.2.2 The Impact of System II Reliability

In the second set of examples, we fix $\mu = 2$ and vary $k_{II} \in \{80, 82, 84, 86, 88, 90\}$ of system II, which is a k_{II}-out-of-100:G system ($n_{II} = 100$) with $\lambda_{II} = 0.009$. Lower k_{II} implies higher reliability for system II. Figure 4 plots T (the maximum A_{II} value below which the HP policy is better than the HF policy) versus k_{II}. Here, we see that lower k_{II} has a similar impact on the performance of the HP policy as higher μ in Figure 3. In order to prefer the HP policy, the difference between A_I and A_{II} does not need to be large if the system II reliability is high. For instance, when $k_{II} = 80$, the HP policy is better in 97 examples out of 100 for $A_{II} < 0.997 = T$.

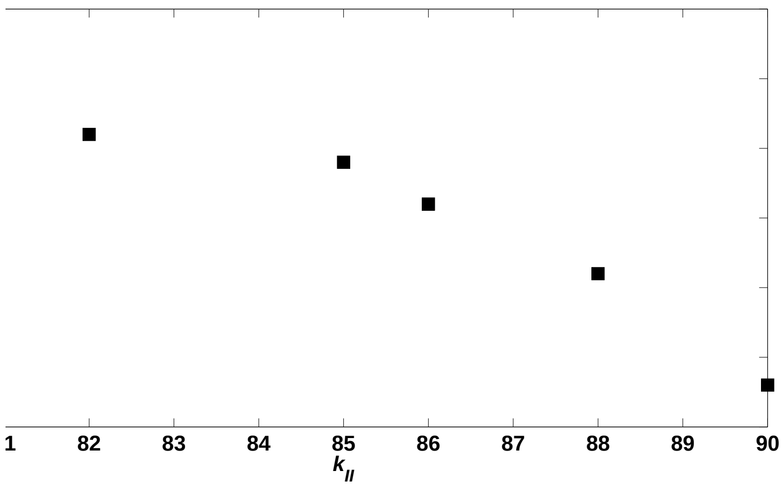

Figure 4: The maximum A_{II} value below which the HP policy outperforms the HF policy vs. k_{II}

As a side note, when k_{II} decreases, the repair shop utilization may increase which, in turn, increases the spare part inventory levels and costs. Figure 5 provides such an example when $A_{II} = 0.95$.

3.2.3 The Impact of System II Failure Rate

In the third set of examples, we fix $\mu = 2$ but vary $n_{II} \in \{20, 50, 70, 80, 90, 100\}$, choosing $k_{II} = \lceil 0.9 n_{II} \rceil$ (greatest integer less than or equal to $0.9 n_{II}$). We set $\lambda_{II} = 0.9/n_{II}$; in other words, the components become less reliable as the size of system II decreases. When $n_{II} = 20$ for which the components are the least reliable, the HF policy always outperforms the HP policy. In this case, the HP policy prioritizes system II for $A_{II} \geq 99.2$, and system I

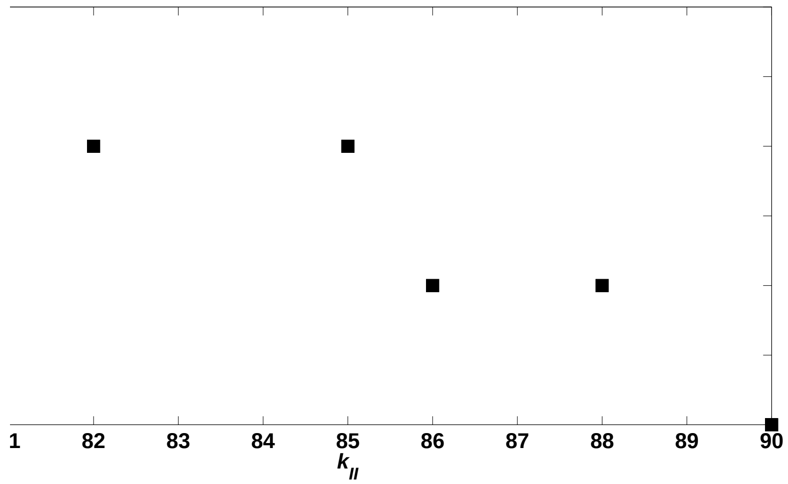

Figure 5: The cost of the HF policy vs. k_{II} when $A_{II} = 0.95$

otherwise. As λ_{II} increases, the HP model tends to give priority to system II. For other n_{II} values, in Figure 6, we plot T (the maximum A_{II} value below which the HP policy is better than the HF policy) versus λ_{II}. Here, we see that if system II has many and more reliable components yielding small λ_{II} values, the difference between A_I and A_{II} does not need to be large in order the HP policy to beat the HF policy. At $n_{II} = 100$, with $\lambda_{II} = 0.009$, the HP policy is better in 78 examples out of 100 for $A_{II} < 0.978 = T$.

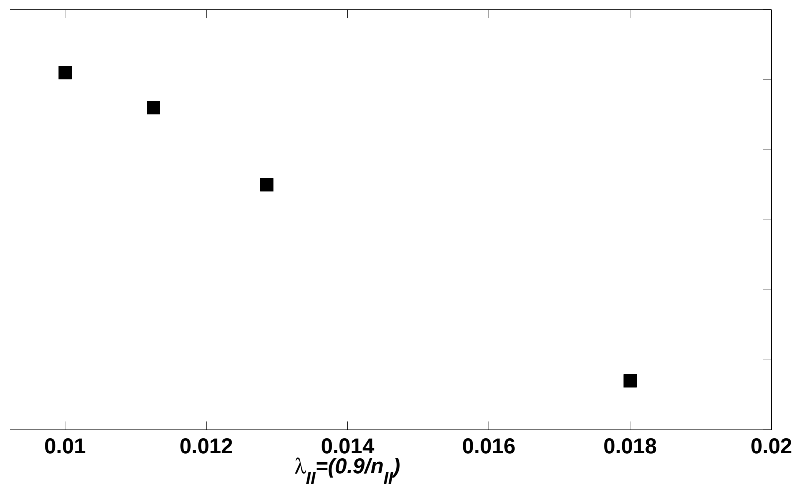

Figure 6: The maximum A_{II} value below which the HP policy outperforms the HF policy vs. λ_{II}

3.2.4 Increasing μ with n_{II}

In the fourth set of examples, we fix $\lambda_{II} = 0.009$ but vary $n_{II} \in \{20, 50, 70, 80, 90, 100\}$, choosing $k_{II} = \lceil 0.9 n_{II} \rceil$. This time, we set $\mu = 1{+}0.01 n_{II}$; in other words, the repair capacity increases with the size of system II. In this set of examples, the HP policy prioritizes system II only when $A_{II} = 0.999$. Figure 7 shows that the behavior of T is not monotone. Only for intermediate values of repair shop capacity (also for medium size system II), do we observe that the targeted availabilities should be wider apart, in order for the HP policy to perform better.

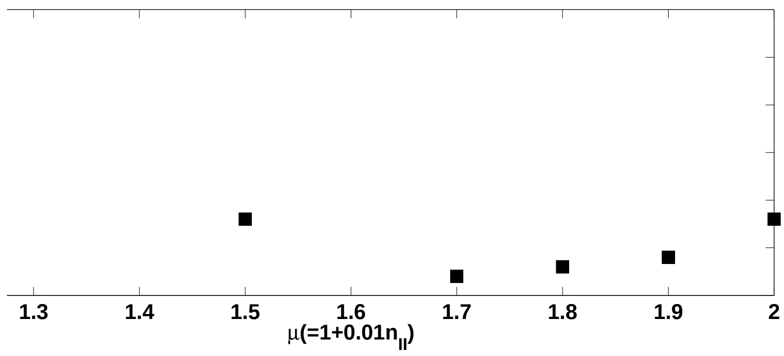

Figure 7: The maximum A_{II} value below which the HP policy outperforms the HF policy vs. $\mu (= 1 + 0.01 n_{II})$

4 Conclusion and Future Research

In this paper, we have analyzed two repair shop/inventory models, namely, the Hybrid-Priority (HP) and the Hybrid-FCFS (HF) models, for a spare part provisioning problem. The arrival rates at the repair shop are state-dependent since the systems served are k-out-of-n:G systems. Our analysis for the FCFS policy can be easily applied to different Markovian settings involving multi-classes of customers with state-dependent arrival rates as long as the service rate is the same for all classes. With our extension, the preemptive-resume priority policy can be also handled when there are more than two classes of customers sharing the same server. The hybrid inventory structure we propose with shared and reserved inventories for each system can have more potential areas of application, as for example, production/inventory systems. Finally, we observe, via our numerical study, that the HP policy is better if the minimum availability expected from each system is not close to the minimum availability of another system.

Acknowledgements

This work was supported in part by Natural Sciences and Engineering Research Council (NSERC) of Canada. The authors thank the two anonymous referees for their invaluable suggestions to improve the manuscript.

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.

Barlow, R. E. and F. Proschan. 1965. *Mathematical theory of reliability*, John Wiley and Sons, New York.

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.

Bitran, G. and R. Caldentey. 2002. "Two-class priority queueing system with state-dependent arrivals", *Queueing Systems*, Vol. 40, **No. 4**, **355–382.**

Boxma, O. J. 1986. "A queueing model of finite and infinite source interaction," *Operations Research Letters* Vol. 5, **No. 5**, 245–254.

Chandra, M. J. 1986. "A Study of multiple finite-source queueing models", *Journal of the Operational Research Society*, Vol. 37, **No. 3**, 275–283.

Dshalalow, J. 1991. "O single-server closed queues with priorities and state dependent parameters", *Queueing Systems*, Vol. 8, **237–253**.

Fawzi, B. B. and A. G. Hawkes. 1991. "Availability of an R-out-of-N system with spares and repairs", *Journal of Applied Probability*, Vol. 28, No. 2, 397–408.

Frostig, E. and B. Levikson. 2002. "On the availability of R out of N repairable systems", *Naval Research Logistics*, Vol. 49, **No. 5**, 483–498.

Graves, S. C. and J. Keilson. 1983. "System balance for extended logistic systems", *Operations Research*, Vol. 31, **No. 2**, 234–252.

Gupta, H. and J. Sharma. 1981. "State transition matrix and transition diagram of k-out-of-n-:G system with spares", *IEEE Transactions on Reliability*, Vol. R-30, No. 4, **395–397**.

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, **No. 2**, 95-105.

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

Kaufman, J. S. 1984. "Finite and Infinite Source Interactions". In: Proceedings of the Tenth International Symposium on Computer Performance Modelling, Measurement and Evaluation. Ed. Erol Gelenbe. *Performance '84*. North-Holland Publishing Co., Amsterdam, The Netherlands, 345–359.

Lawler, E. L. and M. D. Bell. 1966. "A Method for solving discrete optimization problems", *Operations Research*, Vol. 14, **No. 6**, 1098–1112.

Levitin, G. and S. V. Amari. 2010. "Approximation algorithm for evaluating time-to-failure distribution of k-out-of-n system with shared standby elements", *Reliability Engineering and System Safety*, Vol. 95, **No. 4**, 396–401.

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.

Messinger, M. and M. L. Shooman. 1970. "Techniques for optimum spares allocation: A tutorial review", *IEEE Transactions on Reliability*, Vol. R-19, No. 4, 156–166.

Miller, D. R. 1981. "Computation of steady-state probabilities for $M/M/1$ priority queues", *Operations Research*, Vol. 29, **No. 5**, 945–958.

Sahba, P and B. Balcıoğlu. 2011. "The impact of transportation delays on repairshop capacity pooling and spare part inventories", *European Journal of Operational Research*, Vol. 214, **No. 3**, 674–682.

Sasaki, M., S. Kaburaki, and S. Yanagi. 1977. "System availability and optimum spare units", *IEEE Transactions on Reliability*, Vol. 26, No. 3, 182–188.

de Smidt-Destombes, K. S., M. C. van der Heijden, and A. van Harten. 2004. "On the availability of a k-out-of-N system given limited spares and repair capacity under a condition based maintenance strategy", *Reliability Engineering and System Safety*, Vol. 83, **No. 3**, 287–300.

Veran, M. 1984. "Exact analysis of a priority queue with finite source", In: Modelling and Performance Evaluation Methodology. Eds. F. Baccelli and G. Fayolle. *Proceedings of the International Seminar on Modelling and Performance Evaluation Methodology*, **Paris, France, January 24-26, 1983, 371-390.**

Yanagi, S., M. Sasaki, and K. Umazume. 1981. "Optimal inventory problem of a repairable K-out-of-N:G system", *IEEE Transactions on Reliability*, Vol. R-30, No. 5, 478–480.