IISE Transactions

ISSN: 2472-5854 (Print) 2472-5862 (Online) Journal homepage: https://www.tandfonline.com/loi/uiie21

Joint repair sourcing and stocking policies for repairables using Erlang-A and Erlang-B queueing models

Shimon Bitton, Izack Cohen & Morris Cohen

To cite this article: Shimon Bitton, Izack Cohen & Morris Cohen (2019): Joint repair sourcing and stocking policies for repairables using Erlang-A and Erlang-B queueing models, IISE Transactions, DOI: 10.1080/24725854.2018.1560752

To link to this article: https://doi.org/10.1080/24725854.2018.1560752

Accepted author version posted online: 31 Jan 2019.
Published online: 06 May 2019.

IISE TRANSACTIONS
https://doi.org/10.1080/24725854.2018.1560752

Joint repair sourcing and stocking policies for repairables using Erlang-A and Erlang-B queueing models

Shimon Bitton[a], Izack Cohen[a], and Morris Cohen[b]

[a]Industrial Engineering and Management, Technion, Haifa, Israel; [b]Wharton School, University of Pennsylvania, Philadelphia, PA, USA

ABSTRACT

This research focuses on minimizing the life cycle cost of a fleet of aircraft. We consider two categories of repairable parts; upon failure of a first category part (No-Go part), its aircraft becomes non-operational but when a second category part (Go part) fails, the aircraft can still operate for a predetermined period of time before it becomes non-operational. In either case, to minimize aircraft downtime, the failed part has to be replaced with one from the good part inventory, returned from the repair facility or through exchange from a supplier—an emergency sourcing mechanism which is common in the airline industry. Motivated by the observation that a modern aircraft contains a significant fraction of Go parts (estimated at 50% of all repairable parts), we develop a strategic model to decide on stocking and sourcing policies using Erlang-A and Erlang-B queueing models. The suggested model provides an alternative to existing models that typically consider only failed parts that immediately cause a system to be non-operational, and do not consider an emergency sourcing mechanism. A realistic implementation of the model for a fleet of Boeing 737 aircraft, based on a list of 2,805 part types, demonstrates that significant cost savings may be achieved by explicitly modeling Go parts.

ARTICLE HISTORY
Received 20 August 2018
Accepted 4 December 2018

KEYWORDS
Spare parts; stocking; supply chain

1. Introduction

We consider the problem of supporting a fleet of mission-critical systems (our focus is on a fleet of aircraft, e.g., a fleet of Airbus 330s) that are subject to random failures. Fleet availability is achieved by returning failed systems to service promptly. Such a resumption to service is achieved by replacing failed parts with good ones. This is accomplished through the deployment of repair capabilities and the purchase of spare parts, both of which require large capital expenditures. At a more strategic level, firms can also consider options that affect the underlying reliability of aircraft (or systems) in the field and/or the sourcing of parts from alternative (emergency) supply sources.

Minimizing the life cycle costs for such fleets is an important objective in both commercial and defense environments, where capital-intensive equipment must operate reliably over extended periods of ownership and use. The approach most commonly used in practice to solve this problem is based on algorithms that capture the effects of random parts failure, long repair and replenishment lead times, and complex operating environments. The use of such algorithms has become the standard in inventory planning systems currently in-use in industries throughout the world, such as aerospace and defense, semiconductor equipment, automobile and medical equipment.

The research described in this article is motivated by a project conducted with an international airline to improve the management of expensive repairable parts, such as engines, power generators, navigation and communication equipment, etc. The airline conducted a program for the purchase of a new fleet of aircraft, which included large capital expenditures for supporting the fleet through the purchase of expensive repairable parts. We observed that repairable parts in the airline industry, in general, are classified into one of two categories; failures of first category parts cause an immediate grounding of the aircraft until their replacement by good parts (most prior research only considers this category). When parts from the second category fail, the aircraft can still operate for a predetermined period of time, typically 3 to 10 days, after which it is grounded if the failed part is not replaced by a good one. Parts may also be classified into the second category if engineers issue an engineering approval for an aircraft to operate for a limited period of time with a failed part, on a case-by-case basis. It is common to refer to parts in the first category as “No-Go” parts and parts in the second category are referred to as “Go” parts. The fraction of Go parts within a modern aircraft is estimated, by our partner airline, to be half of the overall number of repairable parts. This estimation was confirmed by our examination of a list of 205 repairable parts for a Boeing 737. In this list, Go parts account for 55% of the parts and 50% of the purchase cost. Examples of Go parts (purchase prices, which are tens to 500 thousand dollars per part, are not detailed, due to propriety limitations)

CONTACT Izack Cohen izikc@technion.ac.il
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/uiie.

include electrical power generators, an air data inertial reference unit inside the navigation system, navigation computers, auto pilot system parts, communication equipment, such as radios, entertainment system parts and many more parts, which are spread throughout all aircraft systems.

Supply chain planning models to support system availability in use today typically do not account for Go parts, although it is clear that consideration of such parts could lead to potential savings since they can act as a mechanism to increase the reliability of the system. In other words, unlike existing models, the suggested model accounts for the duration in which a part has failed but the aircraft is still operational.

We also consider an emergency sourcing mechanism, which is commonly named *exchange* by airlines. Typically, an airline in need of a spare part receives it quickly from a third party (hereafter, a supplier) for a fee. The supplier gets the failed part and charges its repair cost to the airline. Today, exchange is done as an emergency response to stockouts. Exchange mechanisms may have different cost structures, such as time-dependent costs, but in this article we consider an expected cost per exchanged item, which is consistent with the literature for emergency procedures (e.g., Van Houtum and Kranenburg, 2015). Henceforth, we refer to such exchanges as an Emergency Procedure (EP) and include the EP within the planning process. As we demonstrate, there are cases in which it is advisable to apply emergency sourcing proactively in order to prevent stockouts instead of doing so reactively, which is typical. For example, one alternative to achieve high availability is to purchase and stock additional parts. Another alternative, which we denote as the proactive EP policy, may be cheaper. Under such a policy, an EP is initiated whenever there is a demand for the last part in stock (rather than when there is a stockout). The suggested model considers such policies and sets the appropriate policy per part type.

Our treatment of the problem is strategic. The objective function is to minimize capital expenditure for parts over the fleet's life-cycle subject to a service level goal based on fleet up-time. Purchase decisions for repairable spare parts (e.g., engines) are typically made at the beginning of a program roll-out phase and are considered to be strategic, since they represent a considerable capital expense and can have a major impact on fleet availability. Since our model explicitly considers emergency sourcing, we imply that there is an incentive for airlines to form strategic emergency sourcing alliances with suppliers in order to get better prices and service performance, and for suppliers to guarantee their performance in a long-term contract. The introduction of a Go part may require re-design of the part and/or changes to its maintenance processes, which are also strategic decisions. Finally we note that the strategic treatment of the problem in this setting is in line with the current trend in the after-sales market towards outsourcing and performance-based contracts (Guajardo *et al.*, 2012).

The main contribution of this article is in developing a repairable parts management and supply chain sourcing model that accounts for the different categories of parts that we have observed in practice. Such a model, which is first of its kind to the best of our knowledge, achieves lower, possibly much lower costs to support a fleet of aircraft. The methodological contributions of this research include the application of Erlang-A and Erlang-B queueing models for modeling Go and No-Go parts, respectively. These queueing models feed into an optimization problem to minimize the sum of life-cycle costs. Our solution approach breaks down the original hard-to-solve optimization problem into a series of efficient solutions, each of which provides the minimal cost for a given downtime value. The solution time for a realistic size problem with thousands of part types is less than 20 minutes on a personal computer.

The problem characteristics that motivated our model can be found in industries other than airlines, where cost-effective after-sales support of complex equipment is a strategic necessity. Our analysis of the underlying parts stocking of both Go and No-Go parts and EP sourcing decisions also leads to insights concerning the strategic use of these options. Thus, this article also contributes to the extensive literature on emergency sourcing, as well as to reliability improvement.

2. A literature survey

The standard model used to quantify the tradeoff between repairable stock levels and system availability is the Metric inventory model (introduced by Feeney and Sherbrooke, 1966; Sherbrooke, 1968). It was originally developed for the United States Air Force for optimizing the deployment of spare parts inventory. The Metric model aims to minimize expected inventory costs subject to meeting a probabilistic service goal based on fleet availability. It does not consider repair costs, as it assumes that all parts have to be repaired and there is no differentiation between repair costs at different locations. Metric-based models have been extended to include multi-indenture problems (Muckstadt, 1973), lateral transshipments (Lee, 1987), cannibalization (Gaver *et al.*, 1993; Sherbrooke, 2004), non-backordering emergency shipments (Cohen *et al.*, 1988) and location-dependent lead times (Wang *et al.*, 2000). Although most of the research considers infinite repair capacity, Diaz and Fu (1997), Perlman *et al.* (2001), and Sleptchenko *et al.* (2002) used a queuing model approximation to prioritize repairs in a finite capacity setting. Sleptchenko *et al.* (2003) also developed a greedy algorithm to decide on capacity and inventory levels. Their modeling approach was extended by Lau and Song (2008) to accommodate non-stationary demands.

A number of papers have considered the joint optimization of repair capacities, inventory investments, and repair sourcing (Alfredsson, 1997; Rappold and van Roo, 2009; Basten *et al.*, 2015; Cohen *et al.*, 2017). Alfredsson (1997) used an Integer Programming (IP) model to decide on repair allocations and Target Stocking Levels (TSLs). Others, such as Rappold and van Roo (2009), used a stochastic integer program to jointly determine repair facility location and inventory allocation. Basten *et al.* (2012) and Basten *et al.* (2015) introduced an IP for a level of repair analysis model of the inventory-allocation problem, which was solved through decomposition and iterative methods. Cohen *et al.* (2017)) considered both central and local (depot) repair policies and developed a heuristic solution algorithm. All of these models are based on a steady state, state-independent policies, which are suitable for longer term planning decisions such as the ones we consider, i.e., how much stock is needed to support a system and what contractual sourcing policies should be used. It is worthwhile to mention models that consider the system's state. Such a model was developed by Caggiano *et al.* (2006) to dynamically decide on repair and inventory allocations for operational day-to-day purposes. These models are typically intractable and use heuristics to find good policies.

There is also a body of research that is directed towards evaluation of the decision to improve the reliability of parts as a means to achieve a desired level of fleet availability. Typically, this decision is considered jointly with the decision to set a stocking level for each part. An economic analysis of this tradeoff was considered in Kim *et al.* (2015) in the context of different service support contracting options, i.e., performance-based or time-and-material, where reliability improvement was based on part failure rates. Another relevant body of research evaluates whether to apply interval-based maintenance or condition-based maintenance (e.g., Christer, 1982). Condition-based maintenance approaches that use information about the actual condition of parts can be used to enhance stocking models and to facilitate improved decisions as demand information becomes available (e.g., advanced demand information research Gallego and Özer, 2001; Howard *et al.*, 2015).

Relaxation/column generation methods represent a solution approach to solve stocking and EP optimization problems (e.g., Kranenburg and van Houtum, 2007; Alvarez *et al.*, 2013). Öner *et al.* (2013) used an Erlang-B queueing model approach for analyzing design decisions where the objective is to minimize life cycle costs of a fleet of systems by deciding on stock, design and repair sourcing policies. The design decision they considered is whether or not to install a redundant part in a system to increase its reliability. Other contributions to modeling reliability improvement can be found in Xie *et al.* (2014) who extended the analysis of redundancy, and in Jin and Tian (2012) who considered a non-stationary version of the problem where the installed base changes over time.

Our model formulation extends Öner *et al.* (2013) to the setting in which an organization does not design the parts used, but instead has to decide on both stocking and EP policies. Unlike previous models in this area, we also consider the two categories of Go and No-Go parts noted above. Also, as noted above, the inclusion of Go parts represents another approach to improving system reliability by allowing for the consequence of a part failure to be delayed in terms of system downtime. To the best of our knowledge, our model is the first to address both EPs and these two part categories simultaneously.

There is an extensive literature on sourcing for repairable parts which includes papers concerned with pooling (lateral transshipments) and emergency sourcing. A recent contribution by Howard *et al.* (2015), incorporated the use of pipeline information to dynamically determine sourcing for cost minimization. Our article contributes to the repairable sourcing literature by considering a setting with availability constraints and by allowing for proactive use of the EP sourcing mechanism, which can be applied in both shortage and non-shortage situations. As noted, we observed its use in emergency (shortage) situations only, in the airline industry and particularly in our partner airline. A related paper by Öner *et al.* (2015) considers the option to proactively replace parts before they fail when a new, more reliable design becomes available.

3. Description of the environment and main assumptions

We examine the problem of minimizing the sum of spare parts purchase costs, inventory holding costs and the source-dependent repair costs, subject to a service constraint based on fleet availability (up-time). We consider a fleet of N identical mission-critical aircraft (e.g., a fleet of Boeing 737s) that has to satisfy a service level goal based on fleet availability. Indeed, airlines and air forces support each fleet of aircraft separately, subject to overall service policies and budget constraints. The model can handle the support of a new fleet of aircraft or the addition of aircraft to an existing fleet (i.e., more failures are expected). In the latter case, we design the model so it can be resolved with the new failure rates; the differences between suggested and previous stocking policies must be adjusted to achieve the desired level of availability. We note that the model introduced in this article can be used in other industries.

We consider a steady state continuous review model, which is suitable for making long-term decisions, such as the purchase of expensive spare parts. The planning horizon, which is 5 to 15 years for our airline partner, is long enough to achieve steady state and its duration, T, may represent the remaining lifetime of the fleet, or a strategic planning period over which the organization plans its expenditures for fleet support. The service level goal corresponds to a fleet downtime D_0 throughout the planning horizon—smaller downtimes correspond to higher fleet availability. A downtime occurs when a part fails and there is a delay in restoring the aircraft to operations, due to parts shortages and the time required to complete a repair. The actual total downtime is the summation of downtime contributions of each part, which amounts to the total time that aircraft are down due to part failures. Consistent with the literature, we assume that parts fail only when aircraft operate (Sherbrooke, 2004; Öner *et al.*, 2013). Time and monetary units are in years and dollars.

Following standard practice, we assume that an aircraft is composed of line replaceable units (parts) $i = 1, \ldots, m$. Each part type i fails, according to a homogeneous Poisson process with rate λ_i. In other words, λ_i is the rate of failures of part type i across the fleet. Indeed, assuming a homogeneous Poisson failure process is reasonable for aerospace and defense environments in which only parts of mature design are used, repairs are performed according to a specification and parts are screened before returning to stock. The short downtimes of systems, relative to the planning horizon T, support the constant failure rate assumption (e.g., Sherbrooke, 2004; Öner *et al.*, 2013; Jin *et al.*, 2015).

These line replaceable units are typically characterized by a high purchase cost, long lead times and have a critical effect on the fleet's availability. They also are typically highly reliable. Thus, an $(s_i - 1, s_i)$ base stock policy is standard, where s_i is the TSL for line replaceable unit i (Sherbrooke, 2004, p. 6). We assume that failures of line replaceable units, within an aircraft, are mutually independent, which is consistent with the literature and with the highly conservative design practices used for aircraft service support.

In this article we only consider line replaceable units and denote them as parts. For simplicity, we assume that only one copy of each type of part is installed within an aircraft (this assumption can be relaxed).

As noted earlier, there are two categories of parts, which we refer to as Go and No-Go. Upon failure of a Go part the aircraft can still operate for a period of time, after which it becomes non-operational. Typically, the "Go duration", G_i, is deterministic, and is defined by regulations or engineering standards. A failed No-Go part causes the aircraft to be immediately non-operational. We model No-Go parts as Go parts with a zero-length "Go duration". Based on analysis of a commercial passenger aircraft, the partner airline that we worked with estimates that the fraction of Go parts is about 50% of all repairables (for example, as noted earlier, based on Boeing data for 2805 Boeing 737 repairable parts, 55% of which are Go). Thus, it is important to distinguish between Go and No-Go parts. Moreover, the fraction of Go parts in other less critical systems may be even larger, and thus our suggested model, based on Go/No-Go classification, is relevant to multiple industries. Our model can also deal with random "Go durations" with an exponential distribution with expected length G_i. The exponential distribution arises in situations where the failed part must first be analyzed in order to determine the specific length of the "Go duration". We would see shorter durations for more complex failures, and longer ones otherwise. Thus, our modeling approach is appropriate for parts that upon failure can still operate, under a specific engineering order, for a limited duration which is dependent upon the nature of the failure and the part.

Regardless of part type, the organization needs to replace failed parts with good ones. Failed parts are sent for repair to the organization's central repair facility, or to a sub-contractor when the organization does not have repair capabilities (both possibilities are considered hereafter as an organizational repair). Repair durations (including transportation times to and from the failure location) for Go and No-Go parts are independent and identically distributed (*i.i.d*) random variables drawn from an exponential and a general distribution, respectively, with an expected value of v_i. We note that while assuming exponentially distributed repair durations is standard in the repairable parts literature, in this article it also follows from the use of Erlang-A $(M/M/s_i + G_i)$ and Erlang-B $(M/G/s_i/s_i)$ models for Go and No-Go parts, respectively (more details are provided in the next section). Repair costs and assembly durations (the time to install a good part in the system which is typically very short, with an order of magnitude of hours), are *i.i.d* random variables from a general distribution with expected values $r_{1,i}$ and $\mu_{1,i}$, respectively. We assume that any part is also available through an EP. Based on the experience of the airline that motivated our model formulation, this assumption is reasonable. So, instead of a repair, the organization can use an EP to replace a failed part with a good one from a supplier. EP durations and costs are *i.i.d* random variables from a general distribution. The expected EP cost is $r_{2,i}$, which includes an EP fee (e.g., for administration) and an expected repair cost. The expected EP duration is $\mu_{2,i}$, which includes the assembly duration. EP costs are assumed to be higher than repair costs for a given part, but have shorter durations. Typically, $r_{2,i} \gg r_{1,i}$ and $\mu_{2,i} \ll \mu_{1,i} + v_i$. Due to the short EP duration (typically, shorter than 48 or 72 hours), we neglect the probability of additional failures of the same part occurring before the part arrives from the EP. Since we are dealing with highly reliable parts, this assumption typically holds. This assumption allows us to consider a single proactive EP policy, as explained in the next section. It can be relaxed at the cost of considering additional policies.

The assumptions, made above, imply that each part type can be treated independently, thus enabling a tractable solution to the multi-part problem.

The independence assumption of part failures within a given system may lead to a conservative downtime calculation in very specific and rare cases. All these cases involve parts failing within the same period in a specific aircraft. Based on real data, the probability for two or more parts failing on a specific aircraft within a Go duration of 3 or 10 days is in the order of magnitude of 10^{-8} or 10^{-4}, respectively. All these events involve combinations of Go parts, and stock-outs for the failed parts; such combinations are rare events for these fleets, which are designed to operate at a high level of availability. In such cases, the downtime of the Go and the other failed part will overlap on the same aircraft, but our model adds them. In light of the small probability for such a sequence of events and the tractable analytical results enabled by the independence assumption, we feel that it is appropriate.

In the aerospace and defense industry, purchase decisions for repairable spare parts (e.g., engines) are typically made at the beginning of a program roll-out phase and are considered to be strategic since they represent a considerable capital expense and can have a major impact on fleet availability. Thus, parts are assumed to be purchased at the beginning of the planning horizon at a unit cost of c_i. Each part has a unit holding cost rate of h_i. We consider an interest rate of α. The objective is to minimize the fleet's total discounted operating costs over the planning horizon. The costs considered are for repair, holding, EP, and for the purchase of spare parts.

Next, we develop a model for Go parts (a single part and multiple parts) followed by a model for No-Go parts. Then, we combine Go and No-Go parts into a unified model that represents an aircraft fleet with airplanes that are composed of both part types.

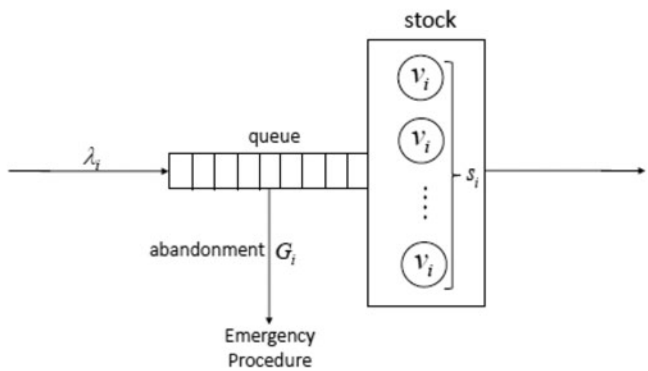

Figure 1. The stock-on-hand process according to an Erlang-A model.

4. Go parts: Analysis, model formulation and solution approach

We first consider Go parts and in the next section formulate the problem for No-Go parts. When a Go part fails there is a "Go duration" during which the aircraft is operational. If there is a part in stock, then it is immediately used to replace the failed part. Otherwise, there is a backorder that can be satisfied by a part that will complete its repair during G_i, or by an EP if there is no such part. If a part arrives from an EP after G_i, there will be some downtime for the aircraft.

We consider two EP policies. According to the first one, in the case of a backorder, we check if a part is expected to return from repair in time, and if not the failed part is sourced to EP. This policy, which is an emergency reactive procedure, is denoted by $z_i = 0$. The second policy is proactive, and is denoted by $z_i = 1$. Under this policy, failed parts are immediately replaced by parts from the stock-on-hand and the only downtime is caused by parts' assembly times. To achieve this, a part is immediately ordered through an EP when the last good part is consumed (that is, the stock-on-hand is one and there is a demand for a good part). Other EP policies, used when the stock-on-hand is larger than one (denoted as $z_i = 2$, $z_i = 3$, etc.) are inferior due to the assumption that no failures occur during the EP duration (following Öner *et al.*, 2013).

4.1. Parts in stock process

We use an Erlang-A model, $M/M/s_i + G_i$ to describe the stochastic process of the stock-on-hand for part i if $z_i = 0$ (see, Figure 1). Failed parts arrive according to a Poisson process with rate λ_i. The number of servers is s_i, which is the TSL, and the service time is exponentially distributed with an expected value of v_i. After service is completed, a part returns to the good inventory (i.e., a server becomes idle) or immediately starts servicing a waiting customer (i.e., a backordered part). If a failed part is not expected to be replaced with a good part during its Go duration, G_i, the EP is used.

The probability to perform an EP for parts is the probability that a failure is not satisfied by a part from stock or repair during G_i, which is the abandonment probability of an Erlang-A model. For a deterministic G_i, the abandonment probability $P(Ab)$ equals (Zeltyn and Mandelbaum, 2005; Mandelbaum and Zeltyn, 2009a):

$$P_{s_i}(Ab) = \frac{1 + \left(\lambda_i - \frac{s_i}{v_i}\right) J_i}{B_i(s_i - 1)^{-1} + \lambda_i J_i}, \tag{1}$$

where

$$B_i(s_i) = \frac{\frac{(\lambda_i v_i)^{s_i}}{s_i!}}{\sum_{j=0}^{s_i} \frac{(\lambda_i v_i)^j}{j!}}, \tag{2}$$

is the Erlang-B blocking probability.

If $\lambda_i - (s_i/v_i) \neq 0$, then

$$J_i = \frac{1}{\frac{s_i}{v_i} - \lambda_i} - \frac{\lambda_i}{\frac{s_i}{v_i}\left(\frac{s_i}{v_i} - \lambda_i\right)} \exp\left(-\left(\frac{s_i}{v_i} - \lambda_i\right) G_i\right)$$

otherwise,

$$J_i = G_i + \frac{v_i}{s_i}. \tag{3}$$

When G_i is the mean of an exponential distribution, then the expression for J_i is:

$$J_i = G_i \exp(\lambda_i G_i) \left(\frac{1}{\lambda_i G_i}\right)^{\frac{s_i G_i}{v_i}} \gamma\left(\frac{s_i G_i}{v_i}, \lambda_i G_i\right),$$

where

$$\gamma(x, y) \equiv \int_0^y t^{x-1} e^{-t} dt, \quad x > 0, \quad y \geq 0,$$

and $\gamma(x, y)$ is an incomplete Gamma function with parameters x, y.

Note that when using $z_i = 1$, failed parts will always be satisfied from stock, i.e., there is no queue (backorders) in such a system and the only downtime is caused by assembly times. For $z_i = 1$, we describe the stochastic parts in stock process using an Erlang-B model, with $s_i - 1$ servers, as indicated in Figure 2. In such case, a failed part is sent for an EP when the system is blocked, where the expected number of EPs during the planning horizon is $\lambda_i T B_i(s_i - 1)$.

In the rest of this article, we consider G_i to be deterministic, since this is the more common situation and also the formulations are simpler. When modeling a random G_i, it is assumed to be exponentially distributed, and then we need to use the relevant formulas for J_i.

4.2. Availability

We represent the fleet's unavailability through its downtime, measured in years. Lower downtimes imply higher availabilities. We denote the downtime goal as D_0. For example, setting $D_0 = 0.1kNT$, where k is the average fraction of the year that the system is operating (e.g., if a system works 24/7 then $k = 1$), corresponds to fleet availability of 90%.

The actual total downtime $D(S, Z)$ is the summation of downtime contributions of each part, $D_i(s_i, z_i)$. The expected number of failures per part i, across the fleet throughout T is λ_{iT}. Each failed part causes its system to be down at least during its assembly duration $\mu_{1,i}$.

The EP arrival time is modeled by the random variable $X_{3,i}$, which is assumed to be exponentially distributed with a mean $\mu_{3,i}$. Note that by definition, $\mu_{3,i} = \mu_{2,i} - \mu_{1,i}$. When modeling Go parts, we need to distinguish between $\mu_{2,i}$ and $\mu_{3,i}$ since if the EP arrival time overlaps the Go duration, the incurred downtime is smaller than the EP arrival time. We follow Baccelli and Hebuterne (1981) who state that the abandon probability does not depend on the time of the abandonment, as long as customers abandon before being serviced. In our context, the equivalent of "abandoning customers" are parts that undergo EP, since they are not expected to enter repair within the Go duration. The probability to abandon depends on repair time, TSL and the length of the Go duration; it is equivalent if abandonments are immediate or postponed. Thus, our policy is to perform EP as soon as possible. We calculate the total expected downtime as follows: For all failed parts the assembly time is incurred. For the abandoning parts, those for which there is a stockout and a good part is not expected to arrive within the Go duration, there is an additional downtime. We note that in real-time management, this policy can be refined by additional considerations based on the actual load in a repair facility and the estimated repair times of specific parts.

Thus, the downtime of part i is:

$$D_i(s_i, z_i) = (1 - z_i) P_{s_i}(Ab) \lambda_i T P(X_{3,i} > G_i) \mathbb{E}[X_{3,i} - G_i]^+ + \lambda_i T \mu_{1,i}. \tag{4}$$

The downtime constraint is:

$$D(S, Z) = \sum_{i=1}^{m} D_i(s_i, z_i) \le D_0. \tag{5}$$

Given the downtime constraint, we can now formulate Problem (P1) for minimizing the overall organization's costs as follows:

$$\begin{aligned} &\min \pi(S, Z) \\ &\text{subject to:} \\ &D(S, Z) \le D_0, \\ &z_i \in \{0, 1\} \quad \forall i \in 1, \ldots, m, \\ &s_i \ge z_i \quad \forall i \in 1, \ldots, m, \\ &s_i \in \mathbb{N}_0 \quad \forall i \in 1, \ldots, m. \end{aligned} \tag{P1}$$

The objective function is a cost function, where,

$$\pi(S, Z) = \sum_{i=1}^{m} \pi_i(s_i, z_i),$$

and the individual cost per part is:

$$\pi_i(s_i, z_i) = S_1(s_i) + S_2(s_i) + R_i(s_i, z_i).$$

The spare parts purchase cost is

$$S_1(s_i) = c_i s_i;$$

the net present value of the holding cost is expressed by

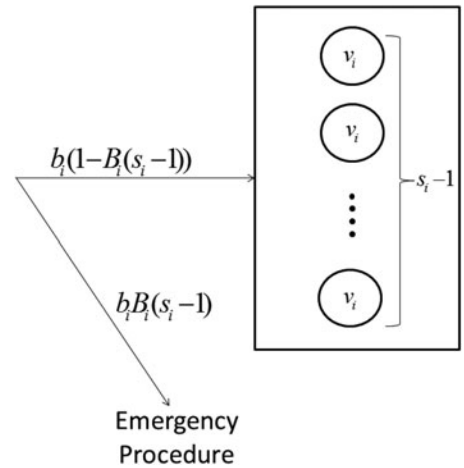

Figure 2. The stock-on-hand process for policy $z_i = 1$.

$$S_2(s_i) = \int_0^T h_i s_i e^{-\alpha T} = \frac{h_i s_i}{\alpha}(1 - e^{-\alpha T}) = s_i T \hat{h}_i,$$

where

$$\hat{h}_i = \frac{h_i}{\alpha T}(1 - e^{-\alpha T}), \tag{6}$$

and the net present value of repair and EP costs are:

$$\begin{aligned} R_i(s_i, z_i) = &\frac{\lambda_i}{\alpha}(1 - e^{-\alpha T})\Big[r_{1,i} + (r_{2,i} - r_{1,i}) P_{s_i}(Ab)(1 - z_i) \\ &+ z_i (r_{2,i} - r_{1,i}) B_i(s_i - 1)\Big]. \end{aligned}$$

We define

$$\hat{r}_{\cdot,i} = \frac{r_{\cdot,i}}{\alpha T}(1 - e^{-\alpha T}), \tag{7}$$

and express $R_i(s_i, z_i)$ more compactly as

$$R_i(s_i, z_i) = \lambda_i T\big[\hat{r}_{1,i} + (\hat{r}_{2,i} - \hat{r}_{1,i})\big[P_{s_i}(Ab)(1 - z_i) + z_i B_i(s_i - 1)\big]\big].$$

Finally,

$$\begin{aligned} \pi_i(s_i, z_i) = &s_i(T\hat{h}_i + c_i) + \lambda_i T\Big[\hat{r}_{1,i} + (\hat{r}_{2,i} - \hat{r}_{1,i})\big[P_{s_i}(Ab)(1 - z_i) \\ &+ z_i B_i(s_i - 1)\big]\Big]. \end{aligned} \tag{8}$$

4.3. Solution approach

Solving Problem (P1) is difficult, since it is a bin packing problem. Thus, we construct an efficient frontier of the cost as a function of the downtime from which the decision maker can choose an appropriate solution to Problem (P1). We use a Lagrangian relaxation approach (Everett, 1963; Fox, 1966), which has been used to solve related problems such as by Öner *et al.* (2013). We solve a bi-objective Problem (P2), which is closely related to Problem (P1); in other words, we find solutions that achieve a minimum cost per a given downtime and a minimum downtime per a given cost:

$$\begin{aligned}&\min \pi(S,Z)\\&\min D(S,Z)\\&\text{subject to:}\\&z_i\in\{0,1\}\quad \forall i\in 1,\dots,m,\\&s_i\geq z_i\quad \forall i\in 1,\dots,m,\\&s_i\in\mathbb{N}_0\quad \forall i\in 1,\dots,m.\end{aligned}\tag{P2}$$

The Lagrangian relaxation approach does not find all of the efficient frontier solutions, but its advantage is in allowing us "to approach a problem first with a simple technique (the basic Lagrange multiplier method) and then to produce additional solutions only when actually desirable or necessary" as Everett (1963) notes. In our settings, when there are several types of parts, the solution approach generates a dense efficient frontier so there is, typically, no need to generate additional solutions, although this is possible. This is demonstrated later in this article through a numerical example. Next, we develop the Lagrangian relaxation approach for the Go parts. A similar approach is used for No-Go parts, with a different Lagrangian formula, and for a combination of both.

The Lagrangian function for Problem P1 is:

$$L(S,Z,\Lambda)=\sum_{i=1}^{m}\pi_i(s_i,z_i)+\Lambda\left(\sum_{i=1}^{m}D_i(s_i,z_i)-D_0\right);$$

$\Lambda\geq 0$ is a unique Lagrange multiplier, which represents the downtime penalty; uniqueness follows from the single downtime constraint.

Denote

$$L_i(s_i,z_i,\Lambda)=\pi_i(s_i,z_i)+\Lambda D_i(s_i,z_i).\tag{9}$$

Thus,

$$L(S,Z,\Lambda)=\sum_{i=1}^{m}L_i(s_i,z_i,\Lambda)-\Lambda D_0.$$

Since we do not use the downtime constraint when constructing the efficient frontier, each part can be considered separately, as shown in Equation (9); In the following sections we consider a single part (Section 4.3.1) and then develop solutions for multiple parts (Section 4.3.2).

4.3.1. Considering a single part

By substituting Equations (4) and (8) in Equation (9) we get:

$$\begin{aligned}L_i(s_i,z_i,\Lambda)=&\ s_i\left(T\hat{h}_i+c_i\right)+\lambda_i T\Big[\hat{r}_{1,i}+(\hat{r}_{2,i}-\hat{r}_{1,i})\\&\left[P_{s_i}(Ab)(1-z_i)+z_iB_i(s_i-1)\right]\Big]\\&+\Lambda\lambda_i T\Big[(1-z_i)P_{s_i}(Ab)P(X_{3,i}>G_i)\mathbb{E}[X_{3,i}-G_i]^+\\&+\mu_{1,i}\Big].\end{aligned}$$

Let $s^*_{i,z_i}(\Lambda)$ be the optimal TSL as a function of Λ for a given z_i policy. We develop our solution following the logic of (Öner *et al.*, 2013) who used an Erlang-B model and the following:

1. Erlang-B blocking probability (Equation (2)) is strictly decreasing and strictly convex in s_i (Karush, 1957).
2. For Go parts, where $z=0$ we conjecture that $P_{s_i}(Ab)$ is convex with respect to s_i. This is justified for several reasons:
   1. The problem of establishing convexity of $P_{s_i}(Ab)$ for $G_i<v_i$ is an open research question, which has been challenging researchers for some time (Koçağa *et al.*, 2015)
   2. The conjecture has been given theoretical support in two ways. One is an actual proof of convexity, under the assumption that $G_i\geq v_i$ (Armony *et al.*, 2009; Koole and Pot, 2011). The other is that $P_{s_i}(Ab)$ converges to the blocking probability of the Erlang-B model, as $G_i\to 0$, and the latter is convex. We follow previous researchers who have made this conjecture (e.g., Mandelbaum and Zeltyn, 2009b, Remark 5.2).
   3. We carried out an extensive numerical experiment with over 127 million separate scenarios to verify that our conjecture is reasonable (see Appendix C). We note that only Go parts with $z_i=0$ need this conjecture since Go parts with $z_i=1$ and No-Go parts do not rely on it.
3. $L_i(s_i,z_i,\Lambda)$ is a linear function of Λ given s_i and z_i.
4. $L_i(s_i,z_i,\Lambda)$ is a convex function of s_i given Λ and z_i (note that $\hat{r}_{2,i}\geq\hat{r}_{1,i}$ and $\mu_{2,i}\geq\mu_{1,i}$; for $s_z=1$, $L_i(s_i,z_i,\Lambda)$ is the sum of convex functions and for $s_z=0$ we also rely on the conjecture about $P_{s_i}(Ab)$ convexity in si).
5. For a given si, $L_i(s_i,0,\Lambda)$ is a linear function of Λ, with a positive slope that decreases with s_i. Under the assumption that $L_i(s_i,0,\Lambda)$ is convex in s_i for a given value of Λ, it follows that $L_i(s^*_{i,0}(0),0,0)\leq L_i(s_i,0,0)$ for every $s_i\geq s^*_{i,0}(0)$. Thus, the function $L_i(s^*_{i,0}(\Lambda),0,\Lambda)$ is a strictly increasing, concave and piecewise linear function of Λ.
6. For a given $s_i\geq 1$ it holds that $L_i(s_i,1,\Lambda)$ is a linear function of Λ.

The intuition for the solution approach is as follows. For every $\Lambda\geq 0$ we find the optimal TSL for $z_i=1$, $s^*_{i,1}(\Lambda)$ and for $z_i=0$, $s^*_{i,0}(\Lambda)$. Thus, the optimal solution for a given Λ is $(s^*_i(\Lambda),z^*_i(\Lambda))=\operatorname{argmin}\{L_i(s_i,z_i,\Lambda)\mid (s_i,z_i)\in\{(s^*_{i,0}(\Lambda),0),(s^*_{i,1}(\Lambda),1)\}\}$. This is used to build an efficient frontier of solutions as explained in the following.

Let $\Delta L_i(s_i,z_i,\Lambda)=L_i(s_i+1,z_i,\Lambda)-L_i(s_i,z_i,\Lambda)$. Under the convexity assumption of $L_i(s_i,0,\Lambda)$ in s_i given Λ and the convexity of $L_i(s_i,1,\Lambda)$ in s_i given Λ it follows that:

$$s^*_{i,z_i}(\Lambda)=\min\{s_i\in\mathbb{N}_0\ \mid\ \Delta L_i(s_i,z_i,\Lambda)\geq 0, s_i\geq z_i\}.\tag{10}$$

We start with policy $z_i=0$. We define $\Delta P_{s_i}(Ab)=P_{s_i}(Ab)-P_{s_i+1}(Ab)$ and develop optimality conditions for TSL and z_i policies (see details in Appendix B),

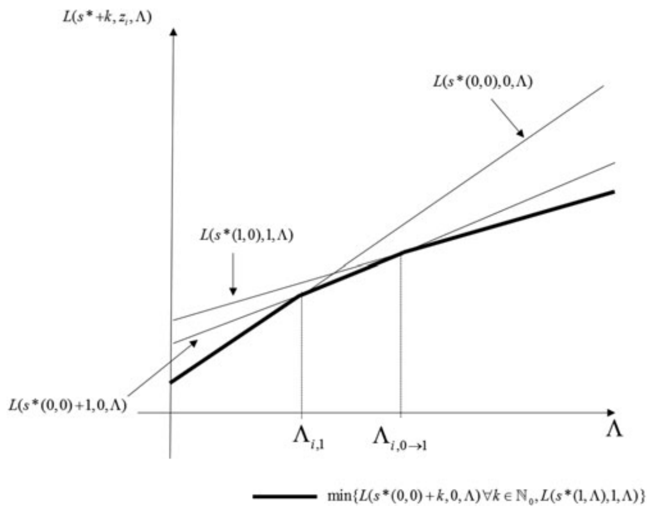

Figure 3. An illustration of the solution process for Go parts.

$$s^*_{i,0}(\Lambda)=\min\left\{s_i\in\mathbb{N}_0|\Delta P_{s_i}(Ab)\right.$$
$$\left.\leq\frac{T\hat{h}_i+c_i}{\lambda_i T\left[(\hat{r}_{2,i}-\hat{r}_{1,i})+\Lambda P(X_{3,i}\ >\ G_i)\mathbb{E}[X_{3,i}-G_i]^+\right]}\right\}.\tag{11}$$

For $z_i=1$ we get

$$s^*_{i,1}(\Lambda)=\min\left\{s_i\in\mathbb{N}|\Delta B_i(s_i-1)\leq\frac{T\hat{h}_i+c_i}{\lambda_i T(\hat{r}_{2,i}-\hat{r}_{1,i})}\right\},\tag{12}$$

which is optimal for all Λ values.

After finding the optimal TSL for every value of Λ and z_i, let us find the solutions $(s^*_i(\Lambda),z^*_i(\Lambda))$ for all values of Λ. We start with $\Lambda=0$ and use Lemma 1, which is proved in Appendix A, to conclude that $z^*_i(0)=0$.

Lemma 1. *For every $s_i\geq 1, P_{s_i}(Ab)<B_i(s_i-1)$.*

Since $P_{s_i}(Ab)<B_i(s_i-1)$, then for every $s_i\in\mathbb{N}, L_i(s_i,0,0)<L_i(s_i,1,0)$ (fewer EPs and the same purchase and holding costs). By definition, $L_i(s^*_{i,0}(0),0,0)\leq L_i(s_i,0,0)$ and $L_i(s^*_{i,1}(0),1,0)\leq L_i(s_i,1,0)$ for every $s_i\in\mathbb{N}$, thus $L_i(s^*_{i,0}(0),0,0)<L_i(s^*_{i,1}(0),1,0)$, which means that for $\Lambda=0$ policy $z_i=0$ outperforms policy $z_i=1$. We now introduce Lemma 2 to find the point of switch from policy $z_i=0$ to policy $z_i=1$ and vice versa (the proof is in Appendix A). In other words, Lemma 2 demonstrates that when high enough availability is needed (i.e., the penalty for downtime is large enough) there exists a $\Lambda_{i,0\to1}>0$ that defines the switch between policies i.e., policy $z_i=1$ outperforms policy $z_i=0$.

Lemma 2. *There exists a $\Lambda_{i,0\to1}>0$ such that for every $\Lambda\geq\Lambda_{i,0\to1}$ policy $z_i=1$ outperforms policy $z_i=0$ and for every $\Lambda<\Lambda_{i,0\to1}$ policy $z_i=0$ outperforms policy $z_i=1$.*

Figure 3 illustrates the suggested solution process. To find $\Lambda_{i,0\to1}$ we first calculate $s^*_{i,0}(0)$. The next step is to find the intersection points that constitute the piecewise linear function discussed above. For example, one needs to find the intersection between $L_i(s^*_{i,0}(0),0,\Lambda)$ and $L_i(s^*_{i,0}(0)+1,0,\Lambda)$, which is defined as $\Lambda_{i,1}$. If this point is above the line $L_i(s^*_{i,1}(\Lambda),1,\Lambda)$ then $\Lambda_{i,0\to1}$ is the intersection point between $L_i(s^*_{i,0}(0),0,\Lambda)$ and $L_i(s^*_{i,1}(\Lambda),1,\Lambda)$, which means that it is optimal to change policy to $z_i=1$ for all $\Lambda\geq\Lambda_{i,0\to1}$. If not, then it is optimal to keep policy $z_i=0$ and $\Lambda_{i,1}$ is the point where it is optimal to add one additional part to stock. Next, we continue by finding the intersection point between $L_i(s^*_{i,0}(0)+1,0,\Lambda)$ and $L_i(s^*_{i,0}(0)+2,0,\Lambda)$. This process continues until $\Lambda_{i,0\to1}$ is found.

Thus, $\Lambda_{i,j}$ are the intersection points where it is optimal to add one stock unit in policy $z_i=0$. In other words, the optimal stock level for Λ where $\Lambda_{i,j-1}\leq\Lambda\leq\Lambda_{i,j}\leq\Lambda_{i,0\to1}$ is $s^*_{i,0}(0)+(j-1)$.

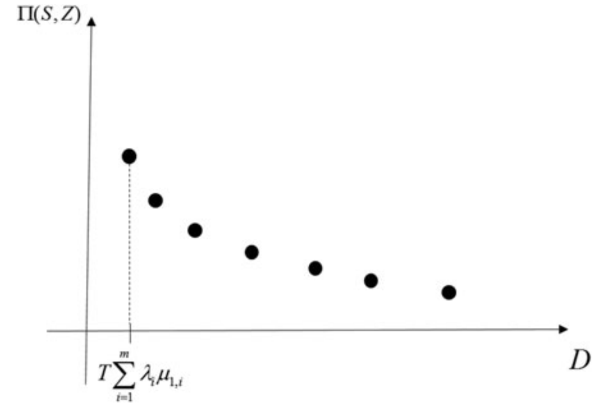

Figure 4. An efficient frontier of the total cost as a function of the downtime for all possible solutions.

4.3.2. Extending the solution to multiple parts

After finding $\Lambda_{i,1},\Lambda_{i,2},...,\Lambda_{i,0\to1}$ for all the parts, the overall fleet's stocking and EP policy is defined as follows: The first step is to construct an efficient frontier of the cost $\pi^*(S,Z)$ as a function of the downtime (see Figure 4). Its first point corresponds to the minimum cost solution given that $\Lambda=0$, $z_i=0$ for all parts and consequently the TSLs correspond to $s^*_{i,0}(0)$ for all parts. Note that this solution yields the largest downtime and it is the cheapest. Next, set $\Lambda=\Lambda'$, find the point which corresponds to $\Lambda=\min\{\Lambda_{i,j},\Lambda_{i,0\to1}|\Lambda_{i,j}>\Lambda',\Lambda_{i,0\to1}>\Lambda',\ \forall i,\ \forall j\}$ and calculate the downtime and cost. Continue until the last solution in which all parts have policy $z_i=1$, the downtime is minimal and the cost is the highest. We note that this approach creates a convex efficient frontier that discretely iterates from one downtime value to smaller ones. There may be other part combinations that may yield intermediate downtime values. To find these combinations, one needs to enumerate all combinations, which is computationally inefficient. Moreover, in reality, when there are several types of parts, the efficient frontier is rather dense (as we demonstrate in the numerical examples) so these intermediate values are not interesting.

5. No-go parts: Analysis, model formulation and solution approach

We consider two EP policies for No-Go parts. According to the first policy, parts are repaired within the organization's

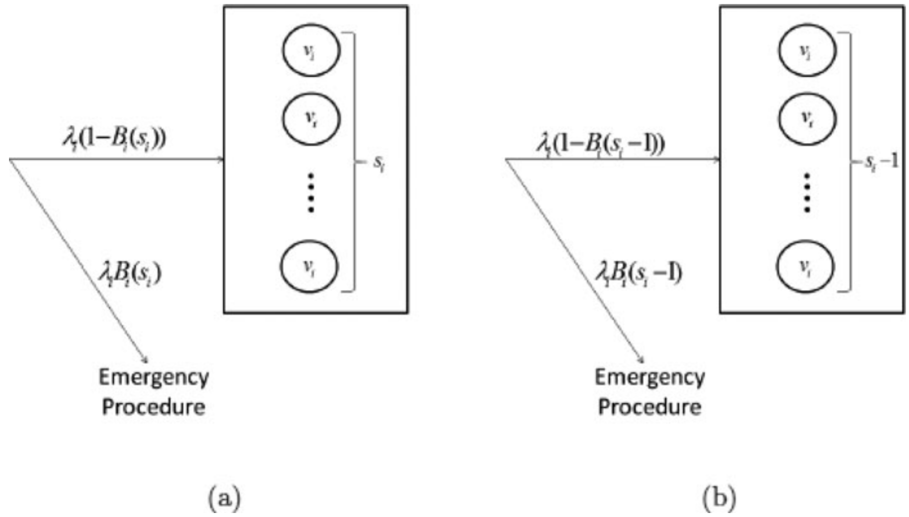

Figure 5. The stock-on-hand process for: (a) Policy $z_i = 0$, and (b) policy $z_i = 1$.

repair facility unless there is a backorder. In such a case, an EP is carried out. This policy is denoted by $z_i = 0$. Under the second policy, denoted by $z_i = 1$, parts are sent to EP when the stock-on-hand is one and there is a demand for a good part. All other EP policies that perform EPs when the stock-on-hand is larger than one are inferior due to the assumption that no failures occur during the EP duration (Öner *et al.*, 2013).

5.1. Parts in stock process

For Go parts, it was natural to use the Erlang-A model to describe the stochastic process of the stock on hand. For No-Go parts, however, if a failed part is not replaced with a good part immediately it is sent for an EP. So, in fact we set the Go duration G_i to 0 and use Lemma 3, which is proved in Appendix A, to find the probability of an EP:

Lemma 3. *When $G_i = 0$ then the probability to abandon an Erlang-A queueing system, $P_{s_i}(Ab)$, is equal to the blocking probability of the corresponding Erlang-B system, $B_i(s_i)$.*

Thus, we use the Erlang-B model, which also holds for $M/G/s_i/s_i$, so the results in the following can be used for general distributed repair durations (Whitt, 2002).

Figure 5 illustrates the stock-on-hand process for policies $z_i = 0$ and $z_i = 1$.

5.2. Availability

We treat the availability in the same manner as for the Go parts. Each failed part causes its system to be down at least during its assembly duration $\mu_{1,i}$. A part from EP causes downtime of $\mu_{2,i}$, which already includes the assembly duration ($\mu_{2,i} = \mu_{3,i} + \mu_{1,i}$). So the expected time of a part's downtime is:

$$D_i(s_i, z_i) = \lambda_i T\left[\mu_{1,i} + (\mu_{2,i} - \mu_{1,i})B_i(s_i)(1 - z_i)\right], \tag{13}$$

and the constraint is:

$$D(S, Z) = \sum_{i=1}^{m} \lambda_i T\left[\mu_{1,i} + (\mu_{2,i} - \mu_{1,i})B_i(s_i)(1 - z_i)\right] \le D_0. \tag{14}$$

5.3. The problem

The optimization problem is identical to Problem P1, but the ingredients are different, as is detailed below (note that $\hat{r}_{\cdot,i}$ and $\hat{h}_{\cdot,i}$ definitions hold). The overall cost is the summation of all parts' costs, $\pi(S, Z) = \sum_{i=1}^{m} \pi_i(s_i, z_i)$, and the cost of each part is:

$$\pi_i(s_i, z_i) = S_1(s_i) + S_2(s_i, z_i) + R_i(s_i, z_i). \tag{15}$$

$S_1(s_i) = c_i s_i$ represents parts' purchase cost at the beginning of the planning horizon, where c_i is the purchase cost per unit.

$S_2(s_i)$ is the net present value of stock holding costs throughout the planning horizon:

$$S_2(s_i) = s_i T\hat{h}_i.$$

The last cost component is the expected net present repair and EP costs:

$$R_i(s_i, z_i) = \lambda_i T\left[\hat{r}_{1,i} + (\hat{r}_{2,i} - \hat{r}_{1,i})\left[(1 - z_i)B_i(s_i) + z_i B_i(s_i - 1)\right]\right].$$

5.4. The model's solution approach

We use the same solution approach used for Go parts; only the Lagrangian formula changes to:

$$\begin{aligned} L_i(s_i, z_i, \Lambda) = {} & s_i\left(T\hat{h}_i + c_i\right) + \lambda_i T \\ & \left[\hat{r}_{1,i} + (\hat{r}_{2,i} - \hat{r}_{1,i})\left[(1 - z_i)B_i(s_i) + z_i B_i(s_i - 1)\right]\right] \\ & + \Lambda\lambda_i T\left[\mu_{1,i} + (\mu_{2,i} - \mu_{1,i})B_i(s_i)(1 - z_i)\right]. \end{aligned}$$

The Lagrangian properties, shown in Section 4.3.1, are valid. Finding $s^*_{i,1}(\Lambda)$ for No-Go parts is identical to Go parts, that is:

$$s^*_{i,1}(\Lambda) = \min\left\{ s_i \in \mathbb{N} | \Delta B_i(s_i - 1) \le \frac{T\hat{h}_i + c_i}{\lambda_i T(\hat{r}_{2,i} - \hat{r}_{1,i})} \right\}, \tag{16}$$

which is optimal for all Λ values.

For $z_i = 0$, we get a different result compared with the Go parts:

$$s^*_{i,0}(\Lambda) = \min\left\{ s_i \in \mathbb{N}_0 | \Delta B_i(s_i) \le \frac{T\hat{h}_i + c_i}{\lambda_i T(\hat{r}_{2,i} - \hat{r}_{1,i}) + \lambda_i T \Lambda(\mu_{2,i} - \mu_{1,i})} \right\}. \tag{17}$$

The rest of the solution approach, for a single part and for multiple parts, follows the procedure described for Go parts, so it is omitted from the text.

6. Combining go and no-go parts

This section starts with a description of the combined solution followed by an illustrative example.

6.1. A combined solution approach

Consider a fleet of aircraft, with m_1 Go and m_2 No-Go parts. The Go parts are numbered $1, .., m_1$ and the No-Go parts $m_1 + 1, ..., m_1 + m_2$. We have shown, for each type of part, how to find the optimal stocking and EP policies for a given downtime penalty. We now formulate the combined problem:

$$\begin{aligned} &\min \sum_{i=1}^{m_1} \pi_i(s_i, z_i) + \sum_{j=m_1+1}^{m_1+m_2} \pi_j(s_j, z_j) \\ &\text{subject to:} \\ &\quad \sum_{i=1}^{m_1} D_i(s_i, z_i) + \sum_{j=m_1+1}^{m_1+m_2} D_j(s_j, z_j) \le D_0, \\ &\quad z_i \in \{0, 1\}, \quad \forall i \in \{1, ..., m_1 + m_2\} \\ &\quad s_i \ge z_i, \quad \forall i \in \{1, ..., m_1 + m_2\} \\ &\quad s_i \in \mathbb{N}_0 \quad \forall i \in \{1, ..., m_1 + m_2\}. \end{aligned} \tag{P3}$$

Note that the $\Lambda \ge 0$ is unique since there is a single availability constraint, and the Lagrangian for the combined problem can be written as:

$$\begin{aligned} L(S, Z, \Lambda) = &\sum_{i=1}^{m_1} \pi_i(s_i, z_i) + \sum_{j=m_1+1}^{m_1+m_2} \pi_j(s_j, z_j) \\ &+ \Lambda\left(\sum_{i=1}^{m_1} D_i(s_i, z_i) + \sum_{j=m_1+1}^{m_1+m_2} D_j(s_j, z_j) - D_0 \right). \end{aligned} \tag{18}$$

Reordering of Equation (18) gives:

$$\begin{aligned} L(S, Z, \Lambda) = &\sum_{i=1}^{m_1} \left(\pi_i(s_i, z_i) + \Lambda D_i(s_i, z_i) \right) \\ &+ \sum_{j=m_1+1}^{m_1+m_2} \left(\pi_j(s_j, z_j) + \Lambda D_j(s_j, z_j) \right) - \Lambda D_0. \end{aligned}$$

Denote

$$L_k(s_k, z_k, \Lambda) = \pi_k(s_k, z_k) + \Lambda D_k(s_k, z_k);$$

thus, $L(S, Z, \Lambda) = \sum_{k=1}^{m_1+m_2} L_k(s_k, z_k, \Lambda) - \Lambda D_0$, which is of identical form as the Lagrangian that was already solved in the previous sections. All the characteristics of the Lagrangian remain intact (e.g., separability, convexity, etc.), so we apply the following solution approach: For each part k we find $\Lambda_{k,1}, \Lambda_{k,2}, ..., \Lambda_{k,0\to 1}$. We iteratively build an efficient frontier following the procedure in Subsection 4.3.2. The final efficient frontier determines the optimal stocking and EP policies for all of the parts for varying values of downtime.

6.2. Numerical examples

We consider two examples to demonstrate the suggested approach: The first example contains five parts, three of which are No-Go and two are Go. Data, provided by the airline with whom we collaborated was used to characterize parts of a Boeing passenger aircraft. The second example illustrates a realistic implementation of the model, for 205 repairable parts. The number of parts was selected based on a list from our partner airline that contained 205 repairable parts of a Boeing 737. For each example we find efficient solutions where each is characterized by a cost and downtime, and determine the EP policy and TSL for all parts. Then, an organization can determine its downtime goal to choose the matching solution.

6.2.1. Example 1

Consider $T = 15$, $m = 5$, and $\alpha = 0.05$. The holding cost is 5% of a unit cost per year.

Assume exponentially distributed EP durations with means $\mu_{3,i}$. Part specific data is outlined in Table 1.

We start by finding the optimal TSL, given $\Lambda = 0$ and $z_i = 0$:

$$s^*_{1,0}(0) = 1, s^*_{2,0}(0) = 2, s^*_{3,0}(0) = 1, s^*_{4,0}(0) = 2, s^*_{5,0}(0) = 3,$$

and $z_i = 1$:

$$s^*_{1,1}(\Lambda) = 2, s^*_{2,1}(\Lambda) = 3, s^*_{3,1}(\Lambda) = 2, s^*_{4,1}(\Lambda) = 3, s^*_{5,1}(\Lambda) = 4.$$

Then, we calculate Λ values for each part, as shown in Table 2.

Table 1. Parts data for Example 1.

	Part 1 (No-Go)	*Part 2 (No-Go)*	*Part 3 (No-Go)*	*Part 4 (Go)*	*Part 5 (Go)*
λ_i	3.6	4.8	2.4	5	6.2
v_i	$\frac{23}{365}$	$\frac{6}{73}$	$\frac{69}{365}$	$\frac{1}{4}$	$\frac{1}{3}$
c_i	465 419	169 355	78 056	50 000	220 000
$r_{1,i}$	14 131	8426	21 650	10 000	25 000
$r_{2,i}$	101 311	43 562	33 846	17 812	59 375
$\mu_{1,i}$	$\frac{1}{4380}$	$\frac{1}{5840}$	$\frac{1}{3504}$	$\frac{1}{2920}$	$\frac{1}{1095}$
$\mu_{2,i}$	$\frac{77}{21\,900}$	$\frac{101}{29\,200}$	$\frac{313}{87\,600}$	$\frac{23}{5475}$	$\frac{53}{14\,600}$
$\mu_{3,i}$	–	–	–	$\frac{6}{1825}$	$\frac{6}{1825}$
G_i	–	–	–	$\frac{3}{365}$	$\frac{10}{365}$
h_i	23 271	8468	3903	2500	11 000

Figure 6 graphically illustrates $L_3(s^*_{i,0}(\Lambda)+k,0,\Lambda)$ for $k=0,1$ and $L_3(s^*_{i,1}(\Lambda),1,\Lambda)$.

The entries in Table 3 specify for each part-solution combination, the policy z_i and the TSL s_i (the last solution, Sol 9, achieves the minimum downtime whereas $z_i = 1 \quad \forall i$). Table 4 presents the costs and downtime for each of the solutions. There are a variety of solutions ranging from Sol 1, which is the cheapest but with the highest downtime, to solutions that contain higher TSLs and the proactive policy of $z_i = 1$ which provides lower downtime at an additional cost. Each one of the nine solutions is optimal in the sense that it provides the highest availability for its cost.

6.2.2. Example 2

This example illustrates a realistic size implementation of the model. We cannot reveal the true data, so we synthesized data that was realistic for the airline industry. We used $T=15$, $m=205$, and $\alpha=0.05$; λ_i was randomly drawn from [2, 34] (based on the MTBF ranges that we observed in the airline industry). The mean repair duration, in years, ν_i was set according to

$$u_i \cdot \frac{1}{3} + \frac{1}{12}$$

where u_i is uniformly distributed between zero and one. Parts cost, c_i was randomly drawn from [50 000, 400 000]; $\mu_{1,i}$ and $\mu_{2,i}$ were set as 6, and 48 hours, respectively. Repair and EP costs, $r_{1,i}$ and $r_{2,i}$ were set to 7% and 17% of c_i, respectively; h_i was set to 5% of c_i. We conducted four different experiments, each with a different percentage of Go parts, 0, 40, 50, and 60%. These percentages were selected based on the list of parts from which 55% are Go and additional estimations provided by our industry partner that about half of the repairable parts within a modern aircraft are Go. The Go duration for these parts was set to 10 days.

Table 2. Λ values for each part.

	$\Lambda_{i,1}$	$\Lambda_{i,0\to 1}$
Part 1	5 710 584	149 265 941
Part 2	16 265 070	50 050 509
Part 3	1 485 934	9 709 310
Part 4	–	33 320 164
Part 5	–	4.374×10^{13}

Table 4. A summary of the downtime and cost for Example 1 solutions.

	Total cost	*Total downtime in years*
Sol 1	7 532 569	0.228
Sol 2	7 575 829	0.199
Sol 3	7 742 464	0.169
Sol 4	7 818 444	0.162
Sol 5	7 995 372	0.151
Sol 6	8 009 148	0.151
Sol 7	8 090 933	0.149
Sol 8	8 635 288	0.145
Sol 9	8 758 123	0.145

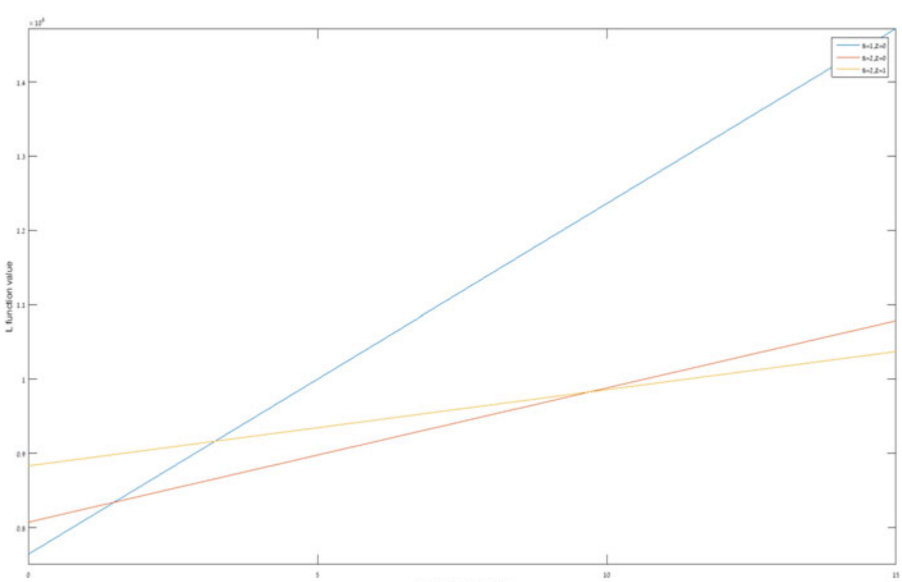

Figure 6. Example 1: An illustration of $L_3(s^*_{i,0}(0)+k,0,\Lambda)$ for $k=0,1$ and $L_3(s^*_{i,1}(\Lambda),1,\Lambda)$.

Table 3. A summary of EP policies and TSLs for Example 1.

	Part 1	*Part 2*	*Part 3*	*Part 4*	*Part 5*
Sol 1	$z_1=0, s_1=1$	$z_2=0, s_2=2$	$z_3=0, s_3=1$	$z_4=0, s_4=2$	$z_5=0, s_5=3$
Sol 2	$z_1=0, s_1=1$	$z_2=0, s_2=2$	$z_3=0, s_3=2$	$z_4=0, s_4=2$	$z_5=0, s_5=3$
Sol 3	$z_1=0, s_1=2$	$z_2=0, s_2=2$	$z_3=0, s_3=2$	$z_4=0, s_4=2$	$z_5=0, s_5=3$
Sol 4	$z_1=0, s_1=2$	$z_2=0, s_2=2$	$z_3=1, s_3=2$	$z_4=0, s_4=2$	$z_5=0, s_5=3$
Sol 5	$z_1=0, s_1=2$	$z_2=0, s_2=3$	$z_3=1, s_3=2$	$z_4=0, s_4=2$	$z_5=0, s_5=3$
Sol 6	$z_1=0, s_1=2$	$z_2=0, s_2=3$	$z_3=1, s_3=2$	$z_4=1, s_4=3$	$z_5=0, s_5=3$
Sol 7	$z_1=0, s_1=2$	$z_2=1, s_2=3$	$z_3=1, s_3=2$	$z_4=1, s_4=3$	$z_5=0, s_5=3$
Sol 8	$z_1=1, s_1=2$	$z_2=1, s_2=3$	$z_3=1, s_3=2$	$z_4=1, s_4=3$	$z_5=0, s_5=3$
Sol 9	$z_1=1, s_1=2$	$z_2=1, s_2=3$	$z_3=1, s_3=2$	$z_4=1, s_4=3$	$z_5=1, s_5=4$

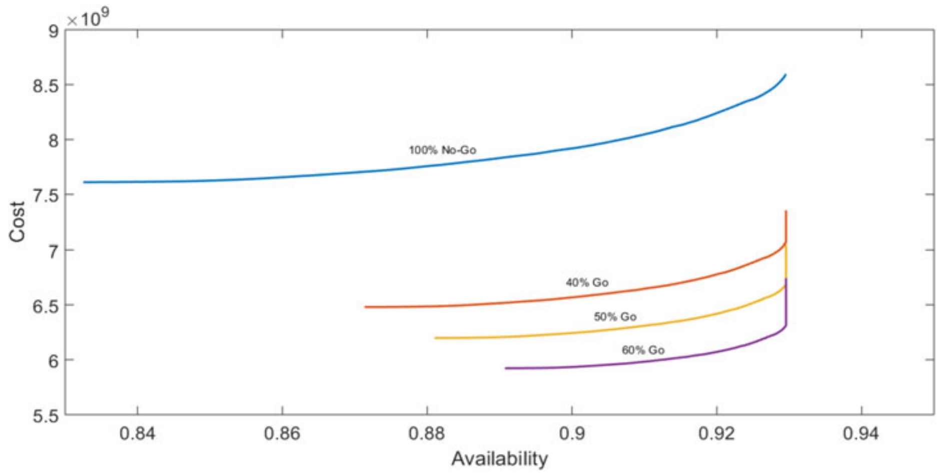

Figure 7. Optimal solution costs as a function of availability for Example 2. The graphs represent 0, 40, 50, and 60% of Go parts out of a total of 2805 parts, respectively.

The solution procedure was programmed in Matlab and the problem was solved. The results, presented in Figure 7, demonstrate a total cost reduction with the number of Go parts. For example 40, 50, and 60% of Go parts reduces the total cost by 18%, 22%, and 26%, respectively. In light of the estimations that the fraction of Go parts in modern commercial aircraft is about 50%, we note that using the suggested model leads to significantly different cost estimates and stocking policies compared to models which do not consider Go parts. To achieve the highest possible availability, all z_is are set to one and the flexibility that stems from having a Go duration is lost.

We note, in passing, that the computational time to run the example for 2805 parts is between 10 to 15 minutes using Matlab and a personal computer with Windows-10 with 16GB RAM, and Intel CPU i7-7500U.

7. Concluding remarks

Our research motivation came from an airline that initiated a purchase program for a new fleet of aircraft; the program included expenditures to support the fleet throughout its life cycle. In this context, we developed a strategic model to manage the inventory of repairable parts in order to minimize life-cycle costs to support a fleet of aircraft that has to satisfy a service level goal based on fleet availability. Unlike previous models that treat all parts as No-Go parts, we consider Go parts using the Erlang-A queueing model. Better decisions can be made when considering Go parts and the flexibility that these parts provide upon failure. This flexibility may lead to a significant reduction in the overall cost. For example, modeling about 50% out of 2805 parts as Go parts, which is in line with estimation of our partner airline for the percentage of Go parts out of all repairable parts, has led to a 22% cost reduction compared with modeling all parts as No-Go. The model also incorporates an EP mechanism that provides flexibility in procuring replacement parts.

The suggested model provides guidelines to organizations, especially those in the airline industry, with respect to purchase decisions of spare parts and the use of EP policies. The model provides an estimation of expected expenditures for EPs. Today, EPs are unplanned reactions to part stockouts, with any supplier that has the needed part. The model introduced here provides incentives for long-term EP agreements with suppliers and the cost estimates provided by our model can be used to support negotiations with suppliers. For example, an organization that forecasts large EP costs for some of its parts can negotiate long-term contracts with suppliers. Under such contracts, the organization may benefit from reduced fees and shorter delivery times (we believe that in some cases the supplier will pre-allocate and position parts close to the organization), which in turn may allow the organization to reduce the quantity of parts purchased. The supplier also may benefit from a long-term relationship with its customers.

Algorithms for solving the model efficiently were developed and resulted in an efficient frontier of solutions, where each solution is characterized by the minimum cost for a specified availability value. Fleet managers can use such results to choose between the efficient solutions—i.e., to explore the tradeoff between cost and fleet availability and select a solution which would be appropriate for their environment.

The suggested model requires additional data compared with standard models, such as length of the Go duration and the EP duration. From our experience with the airline for which this model was developed, only a modest effort is required to gather such additional data and the potential cost savings can be significant.

Notes on contributors

Shimon Bitton is a Ph.D. Student at the Faculty of Industrial Engineering and Management of the Technion. Shimon holds a B.Sc. and an M.E. in Industrial Engineering and Management from the Technion. His current research interests concern distributed and online algorithms.

Dr. Izack Cohen is a faculty member in the Industrial Engineering and Management Faculty at the Technion, Israel. His academic education is from three different Technionic faculties: a B.Sc. in Chemical

Engineering, an M.Sc. in Materials Engineering and a Ph.D. from the Industrial Engineering and Management faculty. Before joining the Technion Izack held senior logistics and engineering management roles in the Israeli Defense Forces. He founded compair (at https://compair.biz), a not-for-profit Israeli initiative that pairs between elderly people and assisted living/home care organizations. Throughout the years Izack consulted to various organizations and promoted various (academic and community-oriented) initiatives.

Dr. Morris A. Cohen is the Panasonic Professor of Manufacturing and Logistics in the Operations, Information and Decisions Department of The Wharton School of the University of Pennsylvania Science. He is also Co-Director of Wharton's Fishman-Davidson Center for Service and Operations Management. He is currently the department editor for services in the Journal of Manufacturing and Service Operations Management and is a Fellow of the Institute for Operations Research and Management Science, and a Senior Fellow of the Manufacturing and Service Operations Management Society. Morris's research interests include analysis of the current drivers of global supply chain sourcing strategy and product-service system modeling with a focus on performance based incentives and buyer-supplier coordination to support a Servicization strategy. He has also developed and implemented advanced optimization tools for supply chain resource planning. His recent application work includes development of strategic and tactical planning systems for service supply chains in industries such as aerospace & defense, consumer electronics, health care technology, oil and gas, automobile, semiconductor equipment, computers, and telecommunications. Until recently Morris was founder and chair of the board of MCA Solutions, a software company specializing in after-sales logistics planning systems, which recently merged with PTC, a leading provider of product design and service life cycle management decision support systems. He recently founded a startup (D3 Analytics), that is applying concepts of machine learning and big data to a new paradigm for supply chain planning and control. He has also been a policy analyst for the planning branch of the Treasury Board Secretariat of the Government of Canada. Morris holds a B.A.Sc. in Engineering Sciences from the University of Toronto, as well as an M.S. in Industrial Engineering and a Ph.D. in Operations Research from Northwestern University.

References

Alfredsson, P. (1997) Optimization of multi-echelon repairable item inventory systems with simultaneous location of repair facilities. *European Journal of Operational Research*, **99**(3), 584–595.

Alvarez, E.M., van der Heijden, M.C. and Zijm, W.H.M. (2013) The selective use of emergency shipments for service-contract differentiation. *International Journal of Production Economics*, **143**(2), 518–526.

Armony, M., Plambeck, E. and Seshadri, S. (2009) Sensitivity of optimal capacity to customer impatience in an unobservable M/M/S queue (Why you shouldn't shout at the DMV). *Manufacturing & Service Operations Management*, **11**(1), 19–32.

Baccelli, F. and Hebuterne, G. (1981) On queues with impatient customers. https://hal.archives-ouvertes.fr/inria-00076467/document.

Basten, R.J.I., van der Heijden, M.C. and Schutten, J.M.J. (2012) Joint optimization of level of repair analysis and spare parts stocks. *European Journal of Operational Research*, **222**(3), 474–483.

Basten, R.J.I., van der Heijden, M.C., Schutten, J.M.J. and Kutanoglu, E. (2015) An approximate approach for the joint problem of level of repair analysis and spare parts stocking. *Annals of Operations Research*, **224**(1), 121–145.

Caggiano, K.E., Muckstadt, J.A. and Rappold, J.A. (2006) Integrated real-time capacity and inventory allocation for reparable service parts in a two-echelon supply system. *Manufacturing & Service Operations Management*, **8**(3), 292–319, 2006.

Christer, A.H. (1982) Modelling inspection policies for building maintenance. *Journal of the Operational Research Society*, **33**(8), 723–732.

Cohen, I., Cohen, M.A. and Landau, E. (2017) On sourcing and stocking policies in a two-echelon, multiple location, repairable parts supply chain. *Journal of the Operational Research Society*, **68**(6), 617–629.

Cohen, M.A., Kleindorfer, P.R. and Lee, H.L. (1988) Service constrained (s, S) inventory systems with priority demand classes and lost sales. *Management Science*, **34**(4), 482–499.

Diaz, A. and Fu, M.C. (1997) Models for multi-echelon repairable item inventory systems with limited repair capacity. *European Journal of Operational Research*, **97**(3), 480–492.

Everett, H. (1963) Generalized Lagrange multiplier method for solving problems of optimum allocation of resources. *Operations Research*, **11**(3), 399–417.

Feeney, G.J. and Sherbrooke, C.C. (1966) The (s-1, s) inventory policy under compound Poisson demand. *Management Science*, **12**(5), 391–411.

Fox, B. (1966) Discrete optimization via marginal analysis. *Management Science*, **13**(3), 210–216.

Gallego, G. and Özer, O. (2001) Integrating replenishment decisions with advance demand information. *Management Science*, **47**(10), 1344–1360.

Gaver, D.P., Isaacson, K.E. and John, B. (1993) Estimating aircraft recoverable spares requirements with cannibalization of designated items, Rand, Technical report, R-4213-AF.

Guajardo, J.A., Cohen, M.A., Kim, S.H. and Netessine, S. (2012) Impact of performance-based contracting on product reliability: An empirical analysis. *Management Science*, **58**(5), 961–979.

Howard, C., Marklund, J., Tan, T. and Reijnen, I. (2015) Inventory control in a spare parts distribution system with emergency stocks and pipeline information. *Manufacturing & Service Operations Management*, **17**(2), 142–156.

Jin, T. and Tian, Y. (2012) Optimizing reliability and service parts logistics for a time-varying installed base. *European Journal of Operational Research*, **218**(1), 152–162.

Jin, T., Xie, W., Liao, H. and Otieno, W. (2015) System availability under redundancy sharing of standby components *Proceedings of the Reliability and Maintainability Annual Symposium* IEEE, Press, Piscataway, NJ, pp.1–6.

Karush, W. (1957) A queuing model for an inventory problem. *Operations Research*, **5**(5), 693–703.

Kim, S., Cohen, M. and Netessine, S. (2015) Reliability or inventory? An analysis of performance-based contracts for product support services, in *Information Exchange in Supply Chain Management*. Cham, Switzerland: Springer.

Koçağa, Y. L., Armony, M. and Ward, A. R. (2015). Staffing call centers with uncertain arrival rates and co-sourcing. *Production and Operations Management*, **24**(7), 1101–1117.

Koole, G. and Pot, A. (2011) Technical note—a note on profit maximization and monotonicity for inbound call centers. *Operations Research*, **59**(5), 1304–1308.

Kranenburg, A.A. and van Houtum, G.J. (2007) Effect of commonality on spare parts provisioning costs for capital goods. *International Journal of Production Economics*, **108**(1), 221–227.

Lau, H.C. and Song, H. (2008) Multi-echelon repairable item inventory system with limited repair capacity under nonstationary demands. *International Journal of Inventory Research*, **1**(1), 67–92.

Lee, H.L. (1987) A multi-echelon inventory model for repairable items with emergency lateral transshipments. *Management Science*, **33**(10), 1302–1316.

Mandelbaum, A. and Zeltyn, S. (2009a) The M/M/n + G queue: Summary of performance measures. http://ie.technion.ac.il/serveng/References/MMNG_formulae.pdf.

Mandelbaum, A. and Zeltyn, S. (2009b) Staffing many-server queues with impatient customers: Constraint satisfaction in call centers. *Operations Research*, **57**(5), 1189–1205.

Muckstadt, J.A. (1973) A model for a multi-item, multi-echelon, multi-indenture inventory system. *Management Science*, **20**(4), 472–481.

Öner, K.B., Scheller-Wolf, A. and van Houtum, G.J. (2013) Redundancy optimization for critical components in high-availability technical systems. *Operations Research*, **61**(1), 244–264.

Öner, K.B., Kiesmüller, G.P. and van Houtum, G.J. (2015) On the upgrading policy after the redesign of a component for reliability improvement. *European Journal of Operational Research*, **244**(3), 867–880.

Perlman, Y. Mehrez, A. and Kaspi, M. (2001) Setting expediting repair policy in a multi-echelon repairable-item inventory system with limited repair capacity. *Journal of the Operational Research Society*, **52**(2), 198–209.

Rappold, J.A. and van Roo, B.D. (2009) Designing multi-echelon service parts networks with finite repair capacity. *European Journal of Operational Research*, **199**(3), 781–792.

Sherbrooke, C.C. (1968) Metric: A multi-echelon technique for recoverable item control. *Operations Research*, **16**(1), 122–141.

Sherbrooke, C.C. (2004) *Optimal Inventory Modeling of Systems: Multi-Echelon Techniques*. Boston, MA: Springer.

Sleptchenko, A., van der Heijden, M.C. and van Harten, A. (2002) Effects of finite repair capacity in multi-echelon, multi-indenture service part supply systems. *International Journal of Production Economics*, **79**(3), 209–230.

Sleptchenko, A., van Der Heijden, M.C. and van Harten, A. (2003) Trade-off between inventory and repair capacity in spare part networks. *Journal of the Operational Research Society*, **54**(3), 263–272.

Van Houtum, G. and Kranenburg, B. (2015) *Spare Parts Inventory Control Under System Availability Constraints*. New York: Springer.

Wang, Y., Cohen, M.A. and Zheng, Y.S. (2000) A two-echelon repairable inventory system with stocking-center-dependent depot replenishment lead times. *Management Science*, **46**(11), 1441–1453.

Whitt, W. (2002) Solutions for the Erlang-B and C formulas. http://www.columbia.edu/~ww2040/ErlangBandCFormulas.pdf.

Xie, W., Liao, H. and Jin, T. (2014) Maximizing system availability through joint decision on component redundancy and spares inventory. *European Journal of Operational Research*, **237**(1), 164–176.

Zeltyn, S. and Mandelbaum, A. (2005) Call centers with impatient customers: Many-server asymptotics of the M/M/n+G queue. *Queueing Systems*, **51**(3-4), 361–402.

Appendix A: Proofs

A.1. Proof of Lemma 1

We use the following: $\lambda_i > 0, v_i > 0, J_i \geq 0$ (see (Mandelbaum and Zeltyn, 2009a) for the definition of J_i), $1 \leq s_i < \infty$ and $B_i(s_i)$ strictly decreasing in s_i (see (Karush, 1957)). We do not provide proof for the case where $J_i = 0$ as it is trivial:

$$P_{s_i}(Ab) = \frac{1+\left(\lambda_i - \frac{s_i}{v_i}\right)J_i}{\frac{1}{B_i(s_i-1)} + \lambda_i J_i} = \frac{1+\left(\lambda_i - \frac{s_i}{v_i}\right)J_i}{\frac{1+\lambda_i J_i B_i(s_i-1)}{B_i(s_i-1)}}$$

$$= \frac{B_i(s_i-1)\left[1+\left(\lambda_i - \frac{s_i}{v_i}\right)J_i\right]}{1+\lambda_i J_i B_i(s_i-1)}$$

$$\frac{B_i(s_i-1)\left[1+\left(\lambda_i - \frac{s_i}{v_i}\right)J_i\right]}{1+\lambda_i J_i B_i(s_i-1)} < B_i(s_i-1)$$

$$\Longleftrightarrow \frac{1+\left(\lambda_i - \frac{s_i}{v_i}\right)J_i}{1+\lambda_i J_i B_i(s_i-1)} < 1$$

$$\Longleftrightarrow 1+\left(\lambda_i - \frac{s_i}{v_i}\right)J_i < 1+\lambda_i J_i B_i(s_i-1)$$

$$\Longleftrightarrow \lambda_i - \frac{s_i}{v_i} < \lambda_i B_i(s_i-1)$$

$$\Longleftrightarrow \lambda_i v_i - s_i < \lambda_i v_i B_i(s_i-1)$$

$$\Longleftrightarrow 1 < B_i(s_i-1) + \frac{s_i}{\lambda_i v_i}$$

$$\Longleftrightarrow 1 < \frac{\frac{(\lambda_i v_i)^{s_i-1}}{(s_i-1)!}}{\sum_{j=0}^{s_i-1}\frac{(\lambda_i v_i)^j}{j!}} + \frac{s_i}{\lambda_i v_i}$$

$$\Longleftrightarrow 1 < \frac{\frac{(\lambda_i v_i)^{s_i-1}}{(s_i-1)!} + \frac{s_i}{\lambda_i v_i}\sum_{j=0}^{s_i-1}\frac{(\lambda_i v_i)^j}{j!}}{\sum_{j=0}^{s_i-1}\frac{(\lambda_i v_i)^j}{j!}}$$

$$\Longleftrightarrow 1 < \frac{\frac{s_i}{\lambda_i v_i}\left[\frac{(\lambda_i v_i)^{s_i}}{(s_i)!} + \sum_{j=0}^{s_i-1}\frac{(\lambda_i v_i)^j}{j!}\right]}{\sum_{j=0}^{s_i-1}\frac{(\lambda_i v_i)^j}{j!}}$$

$$\Longleftrightarrow 1 < \frac{\frac{s_i}{\lambda_i v_i}\sum_{j=0}^{s_i}\frac{(\lambda_i v_i)^j}{j!}}{\sum_{j=0}^{s_i-1}\frac{(\lambda_i v_i)^j}{j!}}$$

$$\Longleftrightarrow 1 < \frac{\frac{s_i}{\lambda_i v_i}\frac{(\lambda_i v_i)^{s_i}}{s_i!}\frac{1}{B_i(s_i)}}{\frac{(\lambda_i v_i)^{s_i-1}}{(s_i-1)!}\frac{1}{B_i(s_i-1)}}$$

$$\Longleftrightarrow 1 < \frac{B_i(s_i-1)}{B_i(s_i)}$$

□

A.2. Proof of Lemma 2

We have shown that function $L_i(s^*_{i,0}(\Lambda), 0, \Lambda)$ is a strictly increasing, concave and piecewise linear function of Λ. For $z_i = 1$, the slope of the line $L_i(s^*_{i,1}(\Lambda), 1, \Lambda)$, $\lambda_i T \mu_{1,i}$, has the smallest possible value—specifically, it is smaller than the slopes of all lines of the piecewise linear function $L_i(s^*_{i,0}(\Lambda), 0, \Lambda)$. Also, $L_i(s^*_{i,0}(0), 0, 0) < L_i(s^*_{i,1}(\Lambda), 1, 0)$. Consequently, there exists a Λ, call it $\Lambda_{i,0\to 1}$ for which $L_i(s^*_{i,0}(\Lambda_{i,0\to 1}), 0, \Lambda_{i,0\to 1}) = L_i(s^*_{i,1}(\Lambda_{i,0\to 1}), 1, \Lambda_{i,0\to 1})$. For all $\Lambda > \Lambda_{i,0\to 1}$ policy $z_i = 1$ outperforms policy $z_i = 0$.

□

A.3. Proof of Lemma 3

We prove for $\lambda_i - (s_i/v_i) \neq 0$. The proof for $\lambda_i - (s_i/v_i) = 0$ is similar. After placing $B(s_1-1)^{-1}$ (Equation (2)) and J_i (Equation (3)) into $P_{s_i}(Ab)$ (Equation (1)) we get:

Table A1. Convexity of $P_s(Ab)$ experiment.

	start value	*end value*	*increment*
$\lambda(\frac{\text{parts}}{\text{year}})$	$\frac{1}{10}$	10	$\frac{1}{10}$
$G(\text{year})$	$\frac{2}{365}$	$\frac{60}{365}$	$\frac{1}{365}$
$v(\text{year})$	$\frac{1}{365}$	$\frac{180}{365}$	$\frac{1}{365}$

$$P_{s_i}(Ab) = \frac{1+\left(\lambda_i-\frac{s_i}{v_i}\right)\left[\frac{1}{\left(\frac{s_i}{v_i}-\lambda_i\right)}-\frac{\lambda_i}{\frac{s_i}{v_i}\left(\frac{s_i}{v_i}-\lambda_i\right)}\right]}{\frac{\sum_{j=0}^{s_i-1}\frac{(\lambda_i v_i)^j}{j!}}{\frac{(\lambda_i v_i)^{s_i-1}}{(s_i-1)!}}+\frac{\lambda_i}{\left(\frac{s_i}{v_i}-\lambda_i\right)}-\frac{\lambda_i^2}{\frac{s_i}{v_i}\left(\frac{s_i}{v_i}-\lambda_i\right)}}$$

$$= \frac{1-\left(\frac{s_i}{v_i}-\lambda_i\right)\left[\frac{1}{\left(\frac{s_i}{v_i}-\lambda_i\right)}-\frac{\lambda_i}{\frac{s_i}{v_i}\left(\frac{s_i}{v_i}-\lambda_i\right)}\right]}{\frac{\sum_{j=0}^{s_i-1}\frac{(\lambda_i v_i)^j}{j!}}{\frac{(\lambda_i v_i)^{s_i-1}}{(s_i-1)!}}+\frac{\lambda_i}{\left(\frac{s_i}{v_i}-\lambda_i\right)}-\frac{\lambda_i^2}{\frac{s_i}{v_i}\left(\frac{s_i}{v_i}-\lambda_i\right)}}$$

$$= \frac{1-\left[1-\frac{\lambda_i v_i}{s_i}\right]}{\frac{\sum_{j=0}^{s_i-1}\frac{(\lambda_i v_i)^j}{j!}}{\frac{(\lambda_i v_i)^{s_i-1}}{(s_i-1)!}}+\frac{\lambda_i}{\left(\frac{s_i}{v_i}-\lambda_i\right)}-\frac{\lambda_i^2}{\frac{s_i}{v_i}\left(\frac{s_i}{v_i}-\lambda_i\right)}}$$

$$= \frac{\frac{\lambda_i v_i}{s_i}}{\frac{\sum_{j=0}^{s_i-1}\frac{(\lambda_i v_i)^j}{j!}}{\frac{(\lambda_i v_i)^{s_i-1}}{(s_i-1)!}}+\frac{\lambda_i}{\left(\frac{s_i}{v_i}-\lambda_i\right)}-\frac{\lambda_i^2}{\frac{s_i}{v_i}\left(\frac{s_i}{v_i}-\lambda_i\right)}}$$

$$= \frac{\lambda_i}{\frac{s_i}{v_i}}\,\frac{1}{\frac{\sum_{j=0}^{s_i-1}\frac{(\lambda_i v_i)^j}{j!}}{\frac{(\lambda_i v_i)^{s_i-1}}{(s_i-1)!}}+\frac{\lambda_i}{\left(\frac{s_i}{v_i}-\lambda_i\right)}-\frac{\lambda_i^2}{\frac{s_i}{v_i}\left(\frac{s_i}{v_i}-\lambda_i\right)}}$$

$$= \frac{\lambda_i}{\frac{\left(\frac{s_i}{v_i}\right)\sum_{j=0}^{s_i-1}\frac{(\lambda_i v_i)^j}{j!}}{\frac{(\lambda_i v_i)^{s_i-1}}{(s_i-1)!}}+\frac{\frac{s_i}{v_i}\lambda_i}{\left(\frac{s_i}{v_i}-\lambda_i\right)}-\frac{\lambda_i^2}{\left(\frac{s_i}{v_i}-\lambda_i\right)}}$$

$$= \frac{\lambda_i}{\frac{\left(\frac{s_i}{v_i}\right)\sum_{j=0}^{s_i-1}\frac{(\lambda_i v_i)^j}{j!}}{\frac{(\lambda_i v_i)^{s_i-1}}{(s_i-1)!}}+\frac{\lambda_i\left(\frac{s_i}{v_i}-\lambda_i\right)}{\left(\frac{s_i}{v_i}-\lambda_i\right)}}$$

$$= \frac{\lambda_i}{\frac{\left(\frac{s_i}{v_i}\right)\sum_{j=0}^{s_i-1}\frac{(\lambda_i v_i)^j}{j!}+\frac{\lambda_i^{s_i}v_i^{s_i-1}}{(s_i-1)!}}{\frac{(\lambda_i v_i)^{s_i-1}}{(s_i-1)!}}}$$

$$= \frac{\frac{\lambda_i^{s_i}v_i^{s_i-1}}{(s_i-1)!}}{\left(\frac{s_i}{v_i}\right)\sum_{j=0}^{s_i-1}\frac{(\lambda_i v_i)^j}{j!}+\frac{\lambda_i^{s_i}v_i^{s_i-1}}{(s_i-1)!}}$$

$$= \frac{\frac{\lambda_i^{s_i}v_i^{s_i-1}}{(s_i-1)!}}{\left(\frac{s_i}{v_i}\right)\sum_{j=0}^{s_i-1}\frac{(\lambda_i v_i)^j}{j!}+\frac{\lambda_i^{s_i}v_i^{s_i-1}}{(s_i-1)!}}\,\frac{\frac{v_i}{s_i}}{\frac{v_i}{s_i}} = B_i(s_i)$$

□

Appendix B: Optimality conditions for TSL and z_i policies: $s^*_{i,0}(\Lambda)$ and $s^*_{i,1}(\Lambda)$

As noted, for a given s_i, $L_i(s_i,0,\Lambda)$ is a linear function of Λ, with a positive slope that decreases with s_i. Under the assumption that $L_i(s_i,0,\Lambda)$ is convex in s_i for a given value of Λ, it follows that $L_i(s^*_{i,0}(0),0,0) \le L_i(s_i,0,0)$ for every $s_i \ge s^*_{i,0}(0)$. Thus, the function $L_i(s^*_{i,0}(\Lambda),0,\Lambda)$ is a strictly increasing, concave and piecewise linear function of Λ. Recall that $\Delta P_{s_i}(Ab) = P_{s_i}(Ab) - P_{s_i+1}(Ab)$. To find $s^*_{i,0}(\Lambda)$ we use the explicit expressions for $L_i(s_i+1,0,\Lambda)$ and $L_i(s_i,0,\Lambda)$ as shown below:

$$\Delta L_i(s_i,0,\Lambda) = L_i(s_i+1,0,\Lambda) - L_i(s_i,0,\Lambda) \ge 0,$$

$$(s_i+1)\left(T\hat{h}_i + c_i\right) + \lambda_i T\left[\hat{r}_{1,i} + (\hat{r}_{2,i}-\hat{r}_{1,i})P_{s_i+1}(Ab)\right]$$

$$+ \Lambda\lambda_i T\Big[P_{s_i+1}(Ab)P(X_{3,i} > G_i)\mathbb{E}[X_{3,i}-G_i]^+ + \mu_{1,i}\Big]$$

$$-[s_i\left(T\hat{h}_i + c_i\right) + \lambda_i T\left[\hat{r}_{1,i} + (\hat{r}_{2,i}-\hat{r}_{1,i})P_{s_i}(Ab)\right]$$

$$+ \Lambda\lambda_i T\Big[P_{s_i}(Ab)P(X_{3,i} > G_i)\mathbb{E}[X_{3,i}-G_i]^+ + \mu_{1,i}\Big]] \ge 0,$$

$$\left(T\hat{h}_i + c_i\right) + \lambda_i T(\hat{r}_{2,i}-\hat{r}_{1,i})\left(P_{s_i+1}(Ab) - P_{s_i}(Ab)\right)$$

$$+ \Lambda\lambda_i TP(X_{3,i} > G_i)\mathbb{E}[X_{3,i}-G_i]^+\left(P_{s_i+1}(Ab) - P_{s_i}(Ab)\right) \ge 0$$

$$\left(T\hat{h}_i + c_i\right) + \lambda_i T(\underbrace{P_{s_i+1}(Ab) - P_{s_i}(Ab)}_{-\Delta P_{s_i}(\mathrm{Ab})})$$

$$[(\hat{r}_{2,i}-\hat{r}_{1,i}) + \Lambda P(X_{3,i} > G_i)\mathbb{E}[X_{3,i}-G_i]^+] \ge 0$$

$$\left(T\hat{h}_i + c_i\right) - \lambda_i T\Delta P_{s_i}(Ab)$$

$$\left[(\hat{r}_{2,i}-\hat{r}_{1,i}) + \Lambda P(X_{3,i} > G_i)\mathbb{E}[X_{3,i}-G_i]^+\right] \ge 0$$

$$\Delta P_{s_i}(Ab) \le \frac{T\hat{h}_i + c_i}{\lambda_i T\left[(\hat{r}_{2,i}-\hat{r}_{1,i}) + \Lambda P(X_{3,i} > G_i)\mathbb{E}[X_{3,i}-G_i]^+\right]};$$

thus,

$$s^*_{i,0}(\Lambda) = \min\Bigg\{s_i \in \mathbb{N}_0 \,|\, \Delta P_{s_i}(Ab) \le \frac{T\hat{h}_i + c_i}{\lambda_i T\left[(\hat{r}_{2,i}-\hat{r}_{1,i}) + \Lambda P(X_{3,i} > G_i)\mathbb{E}[X_{3,i}-G_i]^+\right]}\Bigg\}.$$

In order to find $s^*_{i,1}(\Lambda)$ recall that for a given $s_i \ge 1$ it holds that $L_i(s_i,1,\Lambda)$ is a linear function of Λ. For all $s_i \ge 1$ the slope of $L_i(s_i,1,\Lambda)$ is positive and equals $\lambda_i T\mu_{1,i}$ due to the assumption that no additional parts fail during an EP duration (downtime caused only by the parts' assembly times). Thus, the optimal TSL is independent of Λ. Recalling that $\Delta B_i(s_i) = B_i(s_i) - B_i(s_i+1)$, we use the explicit expressions for $L_i(s_i+1,1,\Lambda)$ and $L_i(s_i,1,\Lambda)$ as shown below to derive $s^*_{i,1}(\Lambda) = s^*_{i,1}(0)$ for all values of Λ:

$$
\begin{aligned}
\Delta L_i(s_i,1,\Lambda) &= L_i(s_i+1,1,\Lambda) - L_i(s_i,1,\Lambda) \geq 0, \\
&(s_i+1)\big(T\hat{h}_i + c_i\big) + \lambda_i T\big[\hat{r}_{1,i} + (\hat{r}_{2,i} - \hat{r}_{1,i})B_i(s_i)\big] \\
&\ + \Lambda\lambda_i T\mu_{1,i} - s_i\big(T\hat{h}_i + c_i\big) \\
&-\lambda_i T\big[\hat{r}_{1,i} + (\hat{r}_{2,i} - \hat{r}_{1,i})B_i(s_i-1)\big] - \Lambda\lambda_i T\mu_{1,i} \geq 0 \\
&T\hat{h}_i + c_i + \lambda_i T(\hat{r}_{2,i} - \hat{r}_{1,i})B_i(s_i) \\
&-\lambda_i T(\hat{r}_{2,i} - \hat{r}_{1,i})B_i(s_i-1) \geq 0, \\
&T\hat{h}_i + c_i - \lambda_i T(\hat{r}_{2,i} - \hat{r}_{1,i})\underbrace{\big(B_i(s_i-1) - B_i(s_i)\big)}_{\Delta B_i(s_i-1)} \geq 0, \\
&-\Delta B_i(s_i-1)\lambda_i T(\hat{r}_{2,i} - \hat{r}_{1,i}) \geq -T\hat{h}_i - c_i, \\
&\Delta B_i(s_i-1) \leq \frac{T\hat{h}_i + c_i}{\lambda_i T(\hat{r}_{2,i} - \hat{r}_{1,i})};
\end{aligned}
$$

thus,

$$
s^*_{i,1}(\Lambda) = \min\left\{ s_i \in \mathbb{N} \middle| \Delta B_i(s_i-1) \leq \frac{T\hat{h}_i + c_i}{\lambda_i T(\hat{r}_{2,i} - \hat{r}_{1,i})} \right\}.
$$

Appendix C: Experiment

We performed an experiment with over 127 000 000 instances, in order to verify that our assumption about convexity of $P_s(Ab)$ in s is reasonable in scenarios which are feasible in the airline industry. The abandon probability is a function of arrival rates, λ, patience, G, expected repair duration, ν and number of servers s.

For each set of values (see Table A1), we checked convexity in s for $s = 1$ to $s = 120$ with increments of one. For every value of s in the test set and for every triplet (λ, G, ν) the function $P_s(Ab)$ was convex.