Multi-item spare parts systems with lateral transshipments and waiting time constraints

Wong, H.; van Houtum, G.J.J.A.N.; Cattrysse, D.; Oudheusden, van, D.

Published: 01/01/2003

Document Version
Publisher’s PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

- A submitted manuscript is the author's version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

Citation for published version (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 22. Oct. 2018
Multi-Item Spare Parts Systems with Lateral Transshipments and Waiting Time Constraints

H. Wong\textsuperscript{a}, G. J. van Houtum\textsuperscript{b}, D. Cattrysse\textsuperscript{a} and D. Van Oudheusden\textsuperscript{a}

\textsuperscript{a} Katholieke Universiteit Leuven, Centre for Industrial Management, Celestijnenlaan 300A, 3001 Heverlee, Belgium
\textsuperscript{b} Technische Universiteit Eindhoven, Faculty of Technology Management, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

Abstract

This paper deals with the analysis of a multi-item, continuous review model of two-location inventory systems for repairable spare parts in which lateral and emergency shipments occur in response to stockouts. A continuous review basestock policy is assumed for the inventory control of the spare parts. The objective is to minimize the total costs for inventory holding, lateral transshipments and emergency shipments subject to a target level for the average waiting time at each of the two locations. A solution procedure based on Lagrangian relaxation is developed to obtain both a lower bound and an upper bound of the optimal total costs. The upper bound follows from a heuristic solution. An extensive numerical experiment shows an average gap of only 0.77\% between the best obtained lower and upper bounds. It also gives insights into the relative improvement achieved when moving from a no-pooling policy to a pooling policy and when moving from an item approach to a system approach. We also applied the model to actual data from an air carrier company.

Keywords: Inventory; Emergency transshipments; Spare parts; Lagrangian relaxation

1. Introduction

Equipment-intensive industries such as airlines, nuclear power plants, various process and manufacturing plants using complex machines are often confronted with the difficult task of maintaining a high system availability, while at the same time there is a pressure to limit the spare parts inventories. A random failure of just one component can cause the system to go
As the downtime can be very costly, spare part inventories are required to keep the downtime to a minimal level. So it is very important to keep the probability of parts being out of stock as low as possible. However, as most parts are quite expensive, maintaining an excessive number of spare parts should also be avoided.

*Lateral transshipments* (also referred to as *inventory pooling*) represent an effective strategy to improve a company’s system availability while reducing the total system costs. Lateral transshipments are used to satisfy a demand at a location that is out of stock from another location with a surplus of on-hand inventory. Since costs for lateral transshipments are generally much lower than downtime costs, lateral transshipments can reduce total system costs. This research was originally motivated by an air carrier company, located in Brussels, who was interested in pooling its spare parts inventories with another company (see Timmers, 1999).

Although a significant amount of research has been done studying various aspects of lateral transshipments in inventory systems, most of it deals with single-item problems in which only one type of part is considered. Such problems are typical when we use an *item approach*. Under an item approach, inventory levels for each individual part are set independently. An alternative approach, denoted as the *system approach* by Sherbrooke (1992a), considers all parts in the system when making inventory-level decisions, and may lead to large reductions in inventory costs in comparison to an item approach (see also Thonemann et al., 2002 and Rustenburg et al., 2003). The main purpose of this paper is to advance the existing literature on multi-item inventory systems with lateral transshipments. Archibald et al. (1997) is the only previous study dealing with such problems. They consider a two-location, multi-item, multi-period, periodic review inventory system with a limited storage space for all items together, i.e. the only connection between the problems for different items is due to the limited storage space. In contrast to their work, we analyze a two-location, multi-item, continuous-review system for repairable items with one-for-one stock replenishments and our optimization problem is to determine stocking policies for all items that minimize the total system costs subject to a target level for the average waiting time for an arbitrary request for a ready-for-use part at each of the two locations. In our model, the decisions with respect to different items are coupled because of the multi-item service measure that is used.

Analyzing lateral transshipments in single-item inventory systems using a continuous review policy with one-for-one stock replenishments is done by Lee (1987), Axšäter (1990), Sherbrooke (1992b), Yanagi and Sasaki (1992), Alfredsson and Verrijdt (1999), Grahovac and Chakravarty (2001), Kukreja et al. (2001), and Wong et al. (2002). One-for-one stock replenishments...
replenishments are common when we deal with (repairable) slow-moving and expensive items. Needham and Evers (1998), Evers (2001), and recently Xu et al. (2003) and Axsäter (2003) considered continuous review \((R, Q)\) policies. Other studies assume periodic review policies and they usually assume no order setup cost, so that an order-up-to or base-stock policy is appropriate. Examples are Gross (1963), Krishnan and Rao (1965), Das (1975), Hoadley and Heyman (1977), Cohen et al. (1986), Karmarkar (1987), Tagaras (1989, 1999), Robinson (1990), Tagaras and Cohen (1992), Archibald et al. (1997), Herer and Rashit (1999), and Rudi et al. (2001) and Herer et al. (2002).

In this paper we also allow emergency supplies from an infinite source when demand at a location cannot be met by either the stock at the local warehouse or the stock at the other locations. This emergency supply mode is also used in Archibald et al. (1997) and Alfredsson and Verrijdt (1999), whereas the other studies consider backlogging. For the systems in which the down time is very costly, e.g. airline companies, the assumption of an emergency supply mode is more realistic than assuming backlogging. Most previous studies do not give exact analysis for the optimization problems with the notable exception of the models of Gross (1963), Krishnan and Rao (1965), Das (1975), Hoadley and Heyman (1977), Robinson (1990), Archibald et al. (1997), and Herer and Rashit (1999). In this paper we derive tight lower and upper bounds of the optimal costs for our multi-item model. The study in this paper can be used as a building block for the analysis of more complex systems. Table 1 presents a brief review of the previous studies and shows the position of this paper in comparison to them.

This paper is organized as follows. In Section 2, we present the problem formulation. We introduce the basic assumptions and the notation of the model, and we present the mathematical formulation our problem. Section 3 describes our solution method which is based on Lagrangian relaxation. We describe how to find the best lower and upper bounds of the optimal objective function. The best lower bound is obtained by optimization of the Lagrange parameters, which is done by the subgradient optimization method. The upper bound follows from a heuristic solution. In Section 4, we perform a computational experiment to show the tightness of the bounds and to study the effect of several parameters on the reduction in costs obtained by applying lateral transshipments and the system approach (in comparison to no lateral transshipments and the item approach). Section 5 presents a model application. We apply our model to actual data from the air carrier company that motivated this work. Finally, we summarize the results in Section 6 and conclude with directions for further research.
<table>
<thead>
<tr>
<th>paper</th>
<th># echelons</th>
<th># items</th>
<th>class of policies</th>
<th>model analysis</th>
<th>important remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gross (1963)</td>
<td>single</td>
<td>single</td>
<td>(s, S)</td>
<td>exact. opt.</td>
<td>• allows the transshipments before the realization of demand</td>
</tr>
<tr>
<td>Krishnan and Rao (1965)</td>
<td>single</td>
<td>single</td>
<td>(s, S)</td>
<td>exact. opt.</td>
<td>• derive optimal solution for two-location systems with identical cost parameters; allow the transshipment after demand is realized but before it must be satisfied</td>
</tr>
<tr>
<td>Das (1975)</td>
<td>single</td>
<td>single</td>
<td>(s, S)</td>
<td>exact. opt.</td>
<td>• extends Gross’s model by permitting transshipments in the middle of each period</td>
</tr>
<tr>
<td>Hoadley and Heyman (1977)</td>
<td>two</td>
<td>single</td>
<td>(s, S)</td>
<td>exact. opt.</td>
<td>• allow balancing acts at the beginning of a period and emergency acts at the end of a period</td>
</tr>
<tr>
<td>Karmarkar (1981)</td>
<td>single</td>
<td>single</td>
<td>general</td>
<td>appr. opt.</td>
<td>• considers the multi-period problem; develops lower and upper bounds of the optimal costs</td>
</tr>
<tr>
<td>Cohen et al. (1986)</td>
<td>multi</td>
<td>single</td>
<td>(s, S)</td>
<td>appr. eval.; appr. opt.</td>
<td>• allow transshipments at higher echelons</td>
</tr>
<tr>
<td>Lee (1987)</td>
<td>two</td>
<td>single</td>
<td>(S-1, S)</td>
<td>appr. eval.; appr. opt.</td>
<td>• assumes identical bases</td>
</tr>
<tr>
<td>Tagaras (1989)</td>
<td>single</td>
<td>single</td>
<td>(s, S)</td>
<td>appr. opt.</td>
<td>• extends the model of Krishnan and Rao (1965) by using a more general cost structure</td>
</tr>
<tr>
<td>Axciater (1990)</td>
<td>two</td>
<td>single</td>
<td>(S-1, S)</td>
<td>appr. eval.</td>
<td>• allows non-identical bases, focus on the demand processes</td>
</tr>
<tr>
<td>Robinson (1990)</td>
<td>single</td>
<td>single</td>
<td>general</td>
<td>exact. opt.</td>
<td>• shows the optimality of a basestock ordering policy when the cost parameters are identical or there are only two locations; analyzes multi-period problems using dynamic programming model</td>
</tr>
<tr>
<td>Tagaras and Cohen (1992)</td>
<td>single</td>
<td>single</td>
<td>(S-1, S)</td>
<td>appr. eval.; appr. opt.</td>
<td>• consider positive replenishment lead times and shows the superiority of complete pooling</td>
</tr>
<tr>
<td>Yanagi and Sasaki (1992)</td>
<td>single</td>
<td>single</td>
<td>(S-1, S)</td>
<td>appr. eval.</td>
<td>• consider limited repair capacity</td>
</tr>
<tr>
<td>Sherbrooke (1992)</td>
<td>two</td>
<td>single</td>
<td>(S-1, S)</td>
<td>appr. eval.</td>
<td>• estimates the expected backorders using a regression model</td>
</tr>
<tr>
<td>Archibald et al. (1997)</td>
<td>single</td>
<td>multi</td>
<td>(s, S)</td>
<td>exact. eval.; exact. opt.</td>
<td>• allow lateral transshipments at any time during a period; model the problem as a Markov decision process</td>
</tr>
<tr>
<td>Needleham and Evers (1998)</td>
<td>two</td>
<td>single</td>
<td>(R, Q)</td>
<td>appr. eval.; appr. opt.</td>
<td>• examine the interaction of relevant costs and transshipment policies; use a simulation-based optimization</td>
</tr>
<tr>
<td>Tagaras (1999)</td>
<td>single</td>
<td>single</td>
<td>(s, S)</td>
<td>appr. eval.</td>
<td>• studies several lateral transshipment policies using a simulation model</td>
</tr>
<tr>
<td>Herer and Rashit (1999)</td>
<td>single</td>
<td>single</td>
<td>general</td>
<td>exact. opt.</td>
<td>• consider fixed and joint replenishment costs; characterize the form of the optimal policy (single period)</td>
</tr>
<tr>
<td>Alfredsson and Verrijdt (1999)</td>
<td>two</td>
<td>single</td>
<td>(S-1, S)</td>
<td>appr. eval.</td>
<td>• allow emergency shipments from the depot and an outside supplier</td>
</tr>
<tr>
<td>Grahovac &amp; Chakravarty (2001)</td>
<td>two</td>
<td>single</td>
<td>(S-1, S)</td>
<td>appr. eval.</td>
<td>• allow lateral transshipments before and after a location is out of stock</td>
</tr>
<tr>
<td>Kukreja et al. (2001)</td>
<td>single</td>
<td>single</td>
<td>(S-1, S)</td>
<td>appr. eval.; appr. opt.</td>
<td>• relax the assumption of exponential repair time distribution</td>
</tr>
<tr>
<td>Evers (2001)</td>
<td>single</td>
<td>single</td>
<td>(R, Q)</td>
<td>appr. opt.</td>
<td>• develops heuristics for determining when transshipments should be made</td>
</tr>
<tr>
<td>Rudi et al. (2001)</td>
<td>single</td>
<td>single</td>
<td>(s, S)</td>
<td>exact. opt.</td>
<td>• consider local decision making; focus on determining transshipment price</td>
</tr>
<tr>
<td>Wong et al. (2002)</td>
<td>single</td>
<td>single</td>
<td>(S-1, S)</td>
<td>appr. eval.; appr. opt.</td>
<td>• consider pooling in a multi-hub system and delayed lateral transshipments</td>
</tr>
<tr>
<td>Herer et al. (2002)</td>
<td>single</td>
<td>single</td>
<td>general</td>
<td>appr. opt.</td>
<td>• prove the optimality of (s,S) policies; consider dynamic transshipments; use a simulation-based optimization</td>
</tr>
<tr>
<td>Xu et al. (2003)</td>
<td>single</td>
<td>single</td>
<td>(R, Q)</td>
<td>appr. eval.</td>
<td>• introduce hold-back parameter which limit the level of outgoing transshipments</td>
</tr>
<tr>
<td>Axsäter (2003)</td>
<td>single</td>
<td>single</td>
<td>(R, Q)</td>
<td>appr. eval.</td>
<td>• allow lateral transshipments only in one direction</td>
</tr>
<tr>
<td>This paper</td>
<td>single</td>
<td>multi</td>
<td>(S-1, S)</td>
<td>exact eval.; appr. opt.</td>
<td>• considers the multi-item problem and employs waiting time constraints</td>
</tr>
</tbody>
</table>
2. Problem formulation

2.1. Notations and assumptions

We model the situation of two independent companies who keep spare parts on stock for their technical systems. Note that throughout this paper the words company and location are used interchangeably. The companies are indexed by \( j = 1, 2 \). We assume that both companies have a number of technical systems of the same type. These systems consist of components which are subject to failures. In total there are \( I \) different items (SKU’s). These items are indexed by \( i = 1, 2, \ldots, I \). Failures occur according to Poisson processes with constant rates. The total failure rate of components of item \( i \) at company \( j \) is given by \( m_{ij} (\geq 0) \). If an item \( i \) does not occur in the configurations of the technical systems at company \( j \), then \( m_{ij} = 0 \). We assume that \( m_{11} + m_{22} > 0 \) for all \( i \). Further, \( M_j = \sum_{i=1}^{I} m_{ij} \) denotes the total failure rate at company \( j \). We assume that \( M_j > 0 \) for \( j = 1, 2 \). Company \( j \) has \( S_j (\in {\mathbb{N}}_0 : = {0} \cup N) \) spare parts of item \( i \). We define \( S_j := (S_j, \ldots, S_j) \). In total, both companies share \( S_i \) spare parts of item \( i \), where \( S_i = S_{1i} + S_{2i} \). All parts are repairable and there is no condemnation. When a part of item \( i \) fails at company \( j \), the failed part is replaced by a spare part. This means that the failed part is immediately removed and sent into repair. A ready-for-use part is put back into the system where the failed part belongs to, as soon as such a part is available. If company \( j \) has a ready-for-use part on stock then this can be done immediately. If not, then there is a waiting time for a ready-for-use part. In that case, if the other company has a ready-for-use part on stock, it sends a part by a lateral transshipment, and the waiting time is limited to the average transshipment time \( ET_{ij}^{\infty} (> 0) \). We assume that complete pooling is applied. This means a company offers its entire available inventory when the other company is experiencing a stockout. If also at the other company no ready-for-use part is available, an emergency supply mode is applied. This means that either the repair operation is expedited or the required part is ordered from an outside supplier e.g., an OEM or a third party supplier. A ready-for-use part becomes available after an average time \( ET_{ij}^{em} (\geq ET_{ij}^{\infty}) \). We believe that for many real life situations, the assumption of an emergency supply mode is more realistic than assuming that one just waits till one of the parts becomes available by the standard repair mode. Failed parts that are sent into repair are returned as ready-for-use parts after exponential repair lead-times. The lead-times of different parts of the same item and of parts of different items are independent. The repair rate of a failed part of item \( i \) is given by \( \mu_i \). We assume that in case a lateral transshipment (an emergency shipment) takes place from company \( j \) (the outside
supplier) to the other company, the failed part will be returned to company \( j \) (the outside supplier) upon completion of its repair. With this assumption, the number of parts on stock plus the number of parts in repair of item \( i \) at company \( j \) is always equal to \( S_{ij} \).

At company \( j \), there is a maximum level \( W_j^{\text{max}} \) given for the average waiting time per request for a ready-for-use part. In this paper, we consider a service model rather than a cost model. In a service model, the objective is to minimize the total system costs subject to a set of service level constraints. In our case, the service level constraints are maximum waiting time constraints. In a cost model, however, the service constraints are replaced with penalty (downtime) costs. Although in general the cost models are analytically more tractable, they have a serious limitation in that the penalty costs are generally hard to quantify. Thus service models are more acceptable from a practical point of view. For a systematic overview of possible relations between the two types of models for general inventory systems, see Van Houtum and Zijm (2000).

Total system costs consist of holding costs, transshipment costs and emergency supply costs. Holding costs \( c^h_i \) are counted for each spare part of item \( i \). A cost \( c^r_i \) is counted for each lateral transshipment of part of item \( i \). A cost \( c^{em}_i \) is counted for each part coming from the emergency supply. The objective is to find a policy \((\overline{S}_1, \overline{S}_2)\) under which the total average costs are minimized subject to the waiting time constraints for the companies 1 and 2.

\[ \sum_{i=1}^{I} \sum_{j=1}^{2} \left( c^h_i x_{ij} + c^r_i \min(x_{ij}, S_{ij}) + c^{em}_i \theta_{ij} \right) \]

To formulate the problem, we define:

\[ \beta_{ij} = \text{fraction of demands for item } i \text{ at company } j \text{ satisfied by its own stocks} \]

\[ \alpha_{ij} = \text{fraction of demands for item } i \text{ at company } j \text{ satisfied by lateral transshipments} \]

\[ \theta_{ij} = \text{fraction of demands for item } i \text{ at company } j \text{ satisfied by emergency supply} \]

\[ W_j = \text{average waiting time per request for a ready-for-use part at company } j \]

Obviously, \( \beta_{ij} + \alpha_{ij} + \theta_{ij} = 1 \) for \( i = 1, \ldots, I; j = 1, 2 \). Since complete pooling is applied here, \( \theta_{ij} \) is the same for \( j = 1, 2 \), i.e. \( \theta_1 = \theta_2 = \theta \) for all \( i \).

The system behavior with respect to an item \( i \) is independent of all other items and may be described by a two-dimensional Markov process. For each item \( i \), we introduce the state \( x_i = (x_{i1}, x_{i2}) \), where \( x_{ij} \) represents the physical stock of spare parts of item \( i \) at company \( j \), and \( 0 \leq x_{ij} \leq S_{ij}, x_{ij} \in \mathbb{N}_0 \). We define \( x_{i1} = (x_{ij}, x_{i2}), x_{i2} = (x_{i1}+1,x_{i2}), x_{i1} = (x_{i1}-1,x_{i2}), x_{i2} = (x_{i1},x_{i2}+1) \).

All possible transitions of the Markov process are as follows:
Transition 1: a failure of a part of item i occurs at location j while $x_{ij} > 0$; the state transition is $x_i \rightarrow x_{ij}$; the transition rate is $m_{ij}$.

Transition 2: a failure of a part of item i occurs at company j while $x_{ij} = 0$ and the other company $j'$ has a positive stock of item i; the state transition is $x_i \rightarrow x_{ij}'$; the transition rate is $m_{ij}$ and this represents a lateral transshipment requested by company j.

Transition 3: a failure of a part of item i occurs at company j while $x_{ij} = x_{ij}' = 0$; an emergency supply is applied; the state transition is $x_i \rightarrow x_i'$; the transition rate is $m_{ij}$.

Transition 4: the repair of a part of item i belonging to company j is completed; the state transition is $x_i \rightarrow x_{ij}'$; the transition rate is $(S_{ij} - x_i)\mu_i$.

Figure 1 shows the Markov process that is obtained when $(S_1, S_2) = (2, 1)$. A similar process is obtained for any other choice for $(S_1, S_2)$.

![Markov process for item i with (S_1, S_2) = (2, 1)](image)

We define $\pi$ as the steady-state probability vector and $\pi_{k,l}$ as the steady-state probability of being in state $(k,l)$, $0 \leq k \leq S_1$, $0 \leq l \leq S_2$. Since the number of states in our problem is not large, a direct method based on Gaussian elimination is applied to determine $\pi$. The fraction of demands for item i satisfied by an emergency supply is equal to the probability of being in state $(0,0)$. Thus, we can write $\theta_i = \pi_{(0,0)}$. This fraction can also be obtained by aggregation on the basis of total physical stock at the locations 1 and 2. That shows that $\theta_i$ is also equal to the Erlang loss probability of an $M/M/S_i/S_i$ queuing system. The fraction of demands for parts of item i at company 1 satisfied by lateral transshipments from company 2 is given by $\alpha_i = \sum_{j=1}^{S_2} \pi_{(i,j)}$. Similarly, $\alpha_i = \sum_{k=1}^{S_2} \pi_{(k,0)}$. The fraction of demands that is satisfied by the local stock is obtained from $\beta_j = 1 - \alpha_j - \theta_i$, $j = 1, 2$. 

---

Note: The image of the Markov process and the diagram are not included in the text here. They are not necessary for understanding the natural text as provided.
We now explain how to obtain the average waiting time. It is possible to aggregate the individual item service (waiting time) functions in several ways. Here, we employ the demand-weighted average service functions as used in Cohen et al. (1992) and Thonemann et al. (2002). For given stocking decisions, the average waiting time per request for a ready-for-use part at company $j$ can be expressed as:

\[
W_j = \sum_{i=1}^{I} \text{Prob}\{\text{an arbitrary failing part at location } j \text{ is of item } i\} \text{ (average waiting time for a ready-for-use part of item } i) \\
= \sum_{i=1}^{I} \frac{m_{ij}}{M_j} (\beta_j 0 + \alpha_j ET_{i}^{\text{pr}} + \theta_j ET_{i}^{\text{em}}) \\
= \sum_{i=1}^{I} \frac{m_{ij}}{M_j} (\alpha_j ET_{i}^{\text{pr}} + \theta_j ET_{i}^{\text{em}}) \tag{1}
\]

With this notation, we can formulate our problem as follows:

**Problem $P_0$:** Minimize 

\[
\sum_{j=1}^{2} \sum_{i=1}^{I} \left( c_i^{\text{h}} S_{ij} + c_i^{\text{p}} m_{ij} \alpha_j + c_i^{\text{em}} m_{ij} \theta_j \right) \tag{2}
\]

subject to \n
\[
\sum_{i=1}^{I} \frac{m_{ij}}{M_j} (\alpha_j ET_{i}^{\text{pr}} + \theta_j ET_{i}^{\text{em}}) \leq W^{\text{max}}_j, \ j = 1, 2, \tag{3}
\]

\[
S_{ij} \in \mathbb{N}_0, \ i = 1, \ldots, I; \ j = 1, 2 \tag{4}
\]

The constraints in (3) can be rewritten as

\[
\sum_{i=1}^{I} m_{ij} (\alpha_j ET_{i}^{\text{pr}} + \theta_j ET_{i}^{\text{em}}) \leq M_j W^{\text{max}}_j, \ j = 1, 2 \tag{5}
\]

**Remark 1** Finding approximations for the fractions of demand satisfied by stock on hand, by lateral transshipments, and by emergency supply has been the focus of previous research; see Lee (1987), Axsäter (1990), Sherbrooke (1992b), Yanagi and Sasaki (1992), Alfredsson and Verrijdt (1999), Grahovac and Chakravarty (2001), Kukreja et al. (2001). Since we want an exact analysis, we work with exact expressions for these fractions instead of approximations.
3. Solution procedure

3.1. The relaxed problem

The problem $P_0$ is an integer-programming problem with a non-linear objective function and non-linear constraints. We apply the Lagrangian relaxation method to solve this problem and derive relations between the relaxation and the original problem in a similar way as in Van Houtum and Zijm (2000). For a given vector $\lambda \in \mathbb{R}^2$ with $\lambda_j \geq 0, j = 1, 2$, we formulate the following problem $P_1$ that is obtained from problem $P_0$ by relaxing the waiting time constraints:

**Problem $P_1$:**

Minimize
\[
\sum_{j=1}^{J} \sum_{i=1}^{I} \left( c_i^s S_{ij} + c_i^m m_{ij} \alpha_{ij} + c_i^e m_{ij} \theta_{ij} \right) + \sum_{j=1}^{J} \lambda_j \left( \sum_{i=1}^{I} n_j (\alpha_j E_{ij}^{sw} + \theta_j E_{ij}^{em}) - M_j W_j^{max} \right)
\]

subject to $S_{ij} \in \mathbb{N}_0, i = 1, \ldots, I; j = 1, 2$

The original problem $P_0$ is a service model, a model in which the objective is to minimize the average total costs subject to the constraints that certain target service levels have to be met. In our case, the target service levels are represented by the maximum waiting time constraints. By putting the service level constraints in the objective function as in the problem $P_1$, we obtain a pure cost model, a model without service level constraints. The pure cost problem can be decomposed into $I$ independent single-item problems.

Let $Z^{\text{B}_0}$ denote the costs of the optimal solution of problem $P_0$, let $Z^{\text{B}_1}(\lambda)$ denote the costs of the optimal solution of problem $P_1$ for given $\lambda = (\lambda_1, \lambda_2)$, and let $Z^{\text{B}_0}(S_1, S_2)$ denote the costs of problem $P_0$ under basestock policy $(S_1, S_2)$. We define $W_j(S_1, S_2)$ as the average waiting time for company $j$ obtained under the stocking policy $(S_1, S_2)$.

**Property 1**

(i) $Z^{\text{B}_0} \geq Z^{\text{B}_1}(\lambda)$ for all $(\lambda_1, \lambda_2) \geq (0,0)$

(ii) $Z^{\text{B}_0} \geq \text{Max}_{\lambda} Z^{\text{B}_1}(\lambda)$
(iii) If for some \( \lambda = (\lambda_1, \lambda_2) \geq (0,0) \) the optimal solution for problem \( P_1 \) is \((S^*_1, S^*_2)\) and \( W_j(S^*_1, S^*_2) \leq W_j^{\max}, \ j = 1, 2, \) then \((S^*_1, S^*_2)\) is feasible for problem \( P_0 \), and
\[
Z^{P_0}(S^*_1, S^*_2) - Z^{P_0} \leq \lambda_1(W_i^{\max} - W_i(S^*_1, S^*_2)) + \lambda_2(W_z^{\max} - W_z(S^*_1, S^*_2)).
\]

(iv) If for some \( \lambda = (\lambda_1, \lambda_2) \geq (0,0) \) the optimal solution for problem \( P_1 \) is \((S^*_1, S^*_2)\) and \( W_j(S^*_1, S^*_2) = W_j^{\max}, \ j = 1, 2, \) then \((S^*_1, S^*_2)\) is the optimal stocking policy for problem \( P_0 \).

**Proof:**

(i) An optimal solution for problem \( P_0 \) is also a feasible solution for problem \( P_1 \) for any \((\lambda_1, \lambda_2) \geq (0,0)\). The costs of this solution for problem \( P_1 \) are smaller than or equal to the costs of this solution for problem \( P_0 \). The costs of an optimal solution of problem \( P_1 \) in their turn are smaller than or equal to the costs of any feasible solution.

(ii) This follows from (i).

(iii) The first part follows from the problem definition in \( P_0 \). The second part follows directly from (i).

(iv) This follows immediately from (iii). ■

By Property 1(i), for any \((\lambda_1, \lambda_2)\) we obtain a lower bound on \( Z^{P_0} \). The maximum value of this lower bound over all considered values \((\lambda_1, \lambda_2)\) is the best obtained lower bound (Property 1(ii)). Besides a lower bound, Property 1 provides us with an upper bound for the distance between the total costs of any feasible solution and the optimal solution (Property 1(iii)). In the next subsection, we describe how to solve problem \( P_1 \) for given values of the Lagrange multipliers.

### 3.2. Solving the relaxed problem for a given \( \lambda \)

The objective function of \( P_1 \) may be rewritten as
\[
\text{Minimize } \sum_{i=1}^{I} \sum_{j=1}^{2} \left( c_i^b S_y + (c_i^p + \lambda_i E_i^p_m) m_i \alpha_i + (c_i^{ew} + \lambda_i E_i ew) m_i \theta_i \right) - \sum_{j=1}^{2} \lambda_j M W_j^{\max} \tag{8}
\]

Now, the \( \lambda_j \) values are assumed to be given. Then the second factor \(-\sum_{j=1}^{2} \lambda_j M W_j^{\max}\) is a constant, and thus can be ignored for optimization purposes. Thus, we are left with \( I \) independent single-item optimization problems. For each item \( i \) the optimization problem can be stated as follows:
Minimize \[ c_i S_i + (c_i^m + \lambda_i T_i^m) m_i \alpha_i + (c_i^m + \lambda_i T_i^m) n_i \theta_i \]
\[ + c_i^h S_i + (c_i^m + \lambda_i T_i^m) m_i \alpha_i + (c_i^m + \lambda_i T_i^m) n_i \theta_i \]  
(9)
subject to \( S_i \in \mathbb{N}_0, S_2 \in \mathbb{N}_0 \)

Recall that \( \alpha_i \) and \( \alpha_2 \) depend on the individual stocks \( S_i \) and \( S_2 \), while \( \theta_i \) depends on the aggregate stocks level \( S \). We can rewrite the objective function as follows:

Minimize \[ f( S_i ) + g( S_i, S_2 ) \]  
(10)
where \[ f( S_i ) = c_i S_i + (c_i^m m_i + \lambda_i T_i^m m_i + c_i^m m_2 + \lambda_i T_i^m m_2 ) \theta_i \]  
(11)
\[ g( S_i, S_2 ) = (c_i^p + \lambda_i T_i^p ) m_i \alpha_i + (c_i^p + \lambda_i T_i^p ) m_2 \alpha_2 \]  
(12)

First, we will look at the behavior of the costs function. It is known that \( \theta_i \) is decreasing and convex as a function of \( S_i \) (see Dowdy et al., 1984); see also Appendix A of Kranenburg and Van Houtum, 2003). As a result, \( f( S_i ) \) is convex. For each value of \( S_i \) there exist \((S_i, S_2)\) with \( S_i + S_2 = S \) such that \( g(S_i, S_2) \) is minimized. Let us define \( g^*(S_i) = \text{Min} \{ g(S_i, S_2) \mid S_i + S_2 = S \} \). For \( S_i = 0 \), no lateral transshipments occur and \( g^*(S_i) = 0 \). Then, increasing \( S_i \) will also increase \( g^*(S_i) \), but when \( S_i \) is large enough, the need for a lateral transshipment diminishes and hence, increasing \( S_i \) will then decrease \( g^*(S_i) \).

We use the following method to obtain the optimal solution. We evaluate the costs over the values of \( S_i \). We start with \( S_i = 0 \) and then increase \( S_i \) incrementally by one. For each \( S_i \), we evaluate the costs \( f(S_i) + g^*(S_i) \), and we keep track of the best solution obtained so far. If we arrive at an \( S_i \) such that \( f(S_i) \) is larger than or equal to the minimum costs obtained so far, we may stop the procedure. It is easy to see that such \( S_i \) value is found in the increasing part of \( f(S_i) \). At this point we can conclude that no better solutions can be found. The best solution obtained so far is an optimal solution. Below we give the optimization procedure in more detail.

**Optimization algorithm for the single-item problem**

**Step 1:** Initialization: Set \( S_i = 0, S_1 = 0, S_2 = 0, \) \( \min = f(0), S_i^* = S_i, S_2^* = S_2 \).

**Step 2:** Set \( S_i = S_i + 1 \) and calculate \( f( S_i ) \). If \( f( S_i ) \geq \min \) go to END, otherwise set \( k = 0 \) and continue.

**Step 3:** Set \( S_i = k \) and \( S_2 = S_i - k \); calculate \( f( S_i ) \) and \( g( S_1, S_2 ) \).
3.3. Finding the tightest lower bound

Once the Lagrangian problem $P_1$ is solved for a given $\lambda$, we know that the objective function $Z^0(\lambda)$ is a lower bound of the optimal costs of the original problem $P_0$. The Lagrange multipliers giving the tightest lower bound are denoted by $\lambda^* = (\lambda_1^*, \lambda_2^*)$. The value of the best possible bound that can be obtained using Lagrangian relaxation is then given by $Z^0(\lambda^*)$ where

$$Z^0(\lambda^*) = \max_{\lambda \geq 0} Z^0(\lambda).$$

(13)

The next step is to find the best Lagrange multipliers $\lambda_1^*$ and $\lambda_2^*$. Since $Z^0(\lambda)$ is not differentiable in general, methods like steepest ascent, which depend on the gradient directions, are not applicable. The subgradient optimization method is usually used instead. It can be viewed as a generalization of the steepest ascent method in which the gradient direction is substituted by a subgradient-based direction (see e.g. Bazarraa et al., 1993). The subgradient optimization is an iterative procedure that has been effective in producing good multiplier values in a variety of Lagrangian-based optimization problems (see Fisher, 1985). We use this method also for solving our problem defined in (13).

Overview of the subgradient optimization method

Let $W_j^k$, $j = 1, 2$, be the expected waiting times that correspond to an optimal solution of the Lagrangian problem for the current vector of multipliers $\lambda^k$ at iteration $k$ of the subgradient method. Then the subgradient direction $\gamma^k = (\gamma_1^k, \gamma_2^k)$ for $\lambda^k$ is calculated as

$$\gamma_j^k = W_j^k - W_j^{\max} \quad \text{for } j = 1, 2.$$  

(14)

At each iteration $k$, the vector of Lagrange multipliers for iteration $(k + 1)$ is updated as
\[ \lambda_j^{k+1} = \max \left( 0, \lambda_j^k - t^k \gamma_j^k \right) \quad \text{for } j = 1, 2. \] (15)

In this formula, \( t^k \) is a scalar stepsize. For the \( t^k \), we use
\[
t^k = s^k \frac{Z^k(\lambda^k) - \hat{Z}}{\|\gamma^k\|}, \tag{16}
\]

which is a stepsize updating procedure that has been used for widespread practical applications. Justification for this formula is given in Held et al. (1974). In this formula, \( \hat{Z} \) is the objective function value of the best known feasible solution to \( P_0 \) (the best upper bound so far) and \( s^k \) is a scalar chosen between 0 and 2. The value of \( s^k \) is halved whenever there is no improvement in the value of the Lagrangian solution after a specified number of iterations.

Usually \( \lambda^0 = 0 \) is the most natural choice as an initial solution. Here, we develop a simple bisection-based procedure to obtain an initial solution that is expected to be close to the optimal solution \( \lambda^* \). We describe the initialization procedure in the following part.

**Initialization procedure**

From Property 1(iii), we know that the optimal vector \( \lambda^* \) will be found at iteration \( k \) where the subgradient direction \( \left( W_1^k - W_1^{\text{max}}, W_2^k - W_2^{\text{max}} \right) \) is close to 0. Our estimation of the location of \( \lambda^* \) is based on the following property.

**Property 2:**
Let \( (S_1^*, S_2^*)_{\lambda_1, \lambda_2} \) denote the optimal solution of problem \( P_1 \) at penalty values \( \lambda_1 \) and \( \lambda_2 \), and \( W_j(S_1^*, S_2^*)_{\lambda_j, \lambda_i}, j = 1, 2 \), the corresponding achieved average waiting times.

(i) For \( \lambda_j^* > \lambda_1 \), \( W_j(S_1^*, S_2^*)_{\lambda_j, \lambda_i} \leq W_j(S_1^*, S_2^*)_{\lambda_1, \lambda_i} \)

(ii) For \( \lambda_2^* > \lambda_2 \), \( W_2(S_1^*, S_2^*)_{\lambda_2, \lambda_j} \leq W_2(S_1^*, S_2^*)_{\lambda_1, \lambda_j} \)


For each value of \( \lambda_2 \) there exists a smallest value of \( \lambda_1 \), say \( \lambda_{1\text{min}} \), at which \( W_1^{\text{max}} \) is satisfied. These points form a function in \( \lambda_2 \). Similarly, we can have \( \lambda_{2\text{min}} \) as a function of \( \lambda_1 \). From Property 2, it follows that of all the points \( (\lambda_1, \lambda_2) \) that give feasible solutions for problem \( P_0 \),
the point where the functions $\lambda_{1\text{min}}(\lambda_2)$ and $\lambda_{2\text{min}}(\lambda_1)$ cross each other (if they do cross) is the best estimate for $\lambda^*$. If both functions do not cross, we would expect that the location of $\lambda^*$ is close to the point $(\lambda_1, \lambda_2)$ where both functions come close to each other. Let us define $(\hat{\lambda}_{1\text{min}}, \hat{\lambda}_{2\text{min}})$ to represent such point.

We performed a computational experiment to learn about the behavior of the functions $\lambda_{1\text{min}}(\lambda_2)$, $\lambda_{2\text{min}}(\lambda_1)$, and the locations of $(\hat{\lambda}_{1\text{min}}, \hat{\lambda}_{2\text{min}})$. Interesting results were obtained from this experiment and several possible situations are depicted in Figure 2(a)-(f). For each situation, we plotted both functions $\lambda_{1\text{min}}(\lambda_2)$ and $\lambda_{2\text{min}}(\lambda_1)$ and also $(\hat{\lambda}_{1\text{min}}, \hat{\lambda}_{2\text{min}})$. In each figure, $(\hat{\lambda}_{1\text{min}}, \hat{\lambda}_{2\text{min}})$ is represented by a small circle. As expected, we found that in general, $\lambda_{1\text{min}}$ is decreasing in $\lambda_2$, and $\lambda_{2\text{min}}$ is decreasing in $\lambda_1$. The locations of $(\hat{\lambda}_{1\text{min}}, \hat{\lambda}_{2\text{min}})$ can be classified as follows:

(a) $\hat{\lambda}_{1\text{min}} = \hat{\lambda}_{2\text{min}}$ (see Figure 2(a) and 2(b)): this occurs when the two companies are identical so that the function $\lambda_{1\text{min}}(\lambda_2)$ and $\lambda_{2\text{min}}(\lambda_1)$ are symmetrical to each other. If the aggregate stocks are even numbers, both companies will have precisely the same number of stocks and both functions $\lambda_{1\text{min}}(\lambda_2)$ and $\lambda_{2\text{min}}(\lambda_1)$ lie on the same line as shown by Figure 1(a).

(b) $\hat{\lambda}_{1\text{min}} < \hat{\lambda}_{2\text{min}}$ (see Figure 2(c) and 2(d)): this occurs in the two following cases: (i) $m_{i1} > m_{i2}$ for all $i$ and $W_{1\text{max}}^* = W_{2\text{max}}^*$; (ii) $m_{i1} \geq m_{i2}$ for all $i$ and $W_{1\text{max}}^* > W_{2\text{max}}^*$. In those two cases, if $\ell$ is an arbitrary value of Lagrange multiplier, we will find that $\lambda_{1\text{min}}(\lambda_2 = \ell) < \lambda_{2\text{min}}(\lambda_1 = \ell)$. In the extreme case, we may find the situation where the first constraint is inactive and $\hat{\lambda}_{1\text{min}} = 0$.

(c) $\hat{\lambda}_{1\text{min}} > \hat{\lambda}_{2\text{min}}$ (see Figure 2(e) and 2(f)): this occurs in the two following cases: (i) $m_{i1} < m_{i2}$ and $W_{1\text{max}}^* = W_{2\text{max}}^*$; (ii) $m_{i1} \leq m_{i2}$ and $W_{1\text{max}}^* < W_{2\text{max}}^*$. By symmetry, the same explanation as under (b) applies here too.

Based on these characteristics, we now describe our initialization procedure. The main idea here is to obtain an initial solution which is expected to be close to the optimal solution of problem (13). In general, we do so by examining three points of $(\lambda_1, \lambda_2)$. For the first point, $\lambda_1$ is set to zero, and we determine smallest value $\lambda_2$ for which both constraints are satisfied. For the second point, we have to find the smallest value $\lambda$, where $\lambda_1 = \lambda_2 = \lambda$, such that both constraints are satisfied. And for the third point, similar to the first point, we set $\lambda_2$ equal to zero and we determine smallest value $\lambda_1$ for which both constraints are satisfied. The point
with the largest objective function value $Z^n(\lambda)$ is then selected as the initial solution for the subgradient optimization procedure. Those three points can be easily determined by applying a standard bisection method. Since the corresponding optimal solution to the selected point is feasible in problem $P_0$, the total costs function $Z^n$ of this solution is used as the initial value for $\hat{Z}$ in the subgradient method.

![Figure 2](image.png)

**Figure 2.** Some possibilities of $\lambda_{1\min}^{1}(\lambda_{2})$, $\lambda_{2\min}^{2}(\lambda_{1})$ and $(\lambda_{1\min}, \lambda_{2\min})$

### 3.4. Finding the best upper bound

In this section we describe the procedure to obtain a good feasible solution for the original problem $P_0$ which provides an upper bound of the optimal total costs. In particular, we are interested in knowing the distance between the best upper bound obtained by this procedure and the best lower bound obtained by the subgradient method described in the previous section.

The procedure works as follows. During the execution of the subgradient method, for each solution $(S_1, S_2)$ that is feasible in problem $P_0$, we evaluate the costs $Z^n(S_1, S_2)$ and we keep track of the best solution obtained so far. If the final solution of the subgradient method is feasible, we stop. If it is not feasible, we apply the procedure described below which may give a better feasible solution than obtained so far.
We consider two ways for obtaining a feasible solution from a non-feasible solution. First, if both constraints are violated, we increase the stock levels in the system. A greedy approach is applied for this purpose. We increase the stock levels for the item that gives the maximum ratio of reduction of waiting time for emergency shipments and extra inventory holding costs $c_i^h$. Next, we put the additional stock at the location where the average waiting time is closest to the target level. This is done until we obtain a feasible solution. Second, if only one of the two constraints is not satisfied, we first try to redistribute the stock at both locations by moving one unit of stock from the location where the constraint is satisfied to the location where the constraint is not satisfied. Since the inventory holding costs are relatively much higher than the transshipment costs, redistributing the stock is usually less costly than increasing the stock levels. The selection of items for the redistribution is done based on the slack parameter (the distance between the maximum waiting time and the individual waiting time) for the location at which the constraint is satisfied. The item which has the largest slack gets the highest priority for the redistribution. This is repeated until a feasible solution is obtained or the two constraints become unsatisfied. In the latter case, we need to proceed with increasing the stock levels. Below we give the procedure in more detail.

**Algorithm for obtaining a feasible solution**

*Input:* A basestock policy $(S_1, S_2)$ with $W_1(S_1, S_2) > W_1^{\text{max}}$ or $W_2(S_1, S_2) > W_2^{\text{max}}$.

*Step 1:* If $W_1(S_1, S_2) > W_1^{\text{max}}$ and $W_2(S_1, S_2) > W_2^{\text{max}}$ go to Step 4, otherwise continue.

*Step 2:* Here we have either (a) $W_1(S_1, S_2) > W_1^{\text{max}}$ and $W_2(S_1, S_2) \leq W_2^{\text{max}}$; or (b) $W_1(S_1, S_2) \leq W_1^{\text{max}}$ and $W_2(S_1, S_2) > W_2^{\text{max}}$. For (a), find item $k$ such that $W_2^{\text{max}} - W_k(S_1, S_2) = \max(W_2^{\text{max}} - W_{k1}(S_1, S_2), ..., W_2^{\text{max}} - W_{k2}(S_1, S_2))$. Set $S_{k2} = S_{k2} - 1$ and $S_{k1} = S_{k1} + 1$. For (b), do the symmetry.

*Step 3:* Calculate $W_{k1}(S_{k11}, S_{k12})$, $W_{k2}(S_{k11}, S_{k12})$, $W_{k1}(S_{k21}, S_{k22})$ and $W_{k2}(S_{k21}, S_{k22})$. If $W_1(S_1, S_2) > W_1^{\text{max}}$ and $W_2(S_1, S_2) > W_2^{\text{max}}$ go to Step 4, otherwise if $W_1(S_1, S_2) \leq W_1^{\text{max}}$ and $W_2(S_1, S_2) \leq W_2^{\text{max}}$ go to END, otherwise go to Step 2.

*Step 4:* For each item $i$, obtain $\rho^i = \frac{(m_{i1} + m_{i2})E T_i^{\text{max}}(\theta_i(S_i) - \theta_i(S_i + 1))}{c_i}$ and choose item $k$ with $\rho^k = \max(\rho^1, ..., \rho^r)$. If $W_1^{\text{max}} - W_1(S_1, S_2) < W_2^{\text{max}} - W_2(S_1, S_2)$ set $S_{k1} = S_{k1} + 1$, otherwise set $S_{k2} = S_{k2} + 1$. Go to Step 3.

END
4. Computational experiment

In this section we present and discuss our numerical findings. Our main focus of inquiry will span:

- the performance of our bounds,
- improvement (in terms of costs) relative to no-pooling solution,
- improvement (in term of costs) relative to the item-approach solution.

Table 3 shows all parameter values for the experiment. All parameter values were selected such that they are realistic for real-life situations, at least for the airline industry. In our experiments the ratio of demand and repair rates for each item were generated randomly from a uniform distribution (notice that only these ratios matter for our problem, and not the precise values for $m_{i1}$, $m_{i2}$ and $\mu_i$). Two uniform distributions were used representing two different variability levels of this ratio among items. From the first distribution, we could have the situation where the maximum value is five times the minimum value while from the second distribution, the ratio of 5 is increased to 30. The same value has been taken for the repair rates of all items, that is, $\mu_i = 0.03$ unit/day. Similarly, the values of the inventory holding costs were generated in the same way. For each combination of $N$, the distribution for generating the $\frac{m_{i1} + m_{i2}}{\mu_i}$, and the distribution for generating the $c_i^k$, ten samples were generated. This results in 120 sample sets. Combined with $2 \times 2 \times 2 = 16$ different possibilities for the other parameters, we obtain 1920 instances in total.

Through our experiments we computed and recorded the following performance measures:

- $\%GAP$ : percentage gap between the upper and the lower bound:
  \[
  \%GAP = \frac{\text{upper bound} - \text{lower bound}}{\text{lower bound}} \times 100
  \]

- $\%SAVE1$ : percentage cost savings when moving from the no-pooling strategy to the pooling strategy. For the problem without lateral transshipments policy, we treated each company independently. For each company, a similar technique using Lagrangian relaxation is used to solve the multi-item optimization problem.

- $\%SAVE2$ : percentage cost savings when moving from the item approach to the system approach. An algorithm similar to the one presented in Table 1 can be used to solve the problem with an item approach. The only difference is that for each item, we need to
check the feasibility of the obtained solutions and the individual waiting time is used instead of the demand-weighted average waiting time.

Table 3. Parameter values for the computational experiments

<table>
<thead>
<tr>
<th>Name of the parameter</th>
<th>Unit</th>
<th>Number of values</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of items (N)</td>
<td></td>
<td>3</td>
<td>20, 50, 100</td>
</tr>
<tr>
<td>Inventory holding costs (c^i_h)</td>
<td>$/unit/year</td>
<td>2</td>
<td>U[5000,15000], U[1000,19000]</td>
</tr>
<tr>
<td>Transshipment costs (c^i_tr)</td>
<td>$</td>
<td>1</td>
<td>250</td>
</tr>
<tr>
<td>Emergency supply costs (c^i_em)</td>
<td>$</td>
<td>2</td>
<td>1250, 2500</td>
</tr>
<tr>
<td>Lateral transshipment lead time (ET^i_tr)</td>
<td>days</td>
<td>2</td>
<td>0.1, 0.25</td>
</tr>
<tr>
<td>Emergency supply lead time (ET^i_em)</td>
<td>days</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Maximum waiting time (W^max = W^max)</td>
<td>days</td>
<td>2</td>
<td>0.25, 0.1</td>
</tr>
<tr>
<td>Demand rates (m_i + m_o)</td>
<td></td>
<td>2</td>
<td>U[0.5,2.5], U[0.1,3.0]</td>
</tr>
<tr>
<td>Repair rates (1/\mu_i)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Demand rates at company 1 (m_i)</td>
<td></td>
<td>2</td>
<td>1, 3</td>
</tr>
<tr>
<td>Demand rates at company 2 (m_o)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The results of our experiments are presented in Table 4(a)-(h). Each table records the average values for: the total costs of the best feasible solutions, %GAP, %SAVE1, and %SAVE2. In Table 4(a)-(g) we show how the performance measures behave with respect to the difference of several parameters. The average results for all instances are presented in Table 4(h). The main observations drawn from Table 4(a)-(h) are summarized as follows:

- For the parameter set used in this experiment, we observe that the performance of our bounds is very good as indicated by very low %GAP (with the average of 0.77%). We also observe that %GAP is decreasing in N. Such behavior is common for many optimization problems with integer decision variables, e.g. Knapsack problems, where the heuristic can provide a reasonable approximation if the number of items is large enough. The %GAP is rather insensitive for the other parameters.

- The average percentage cost savings gained from allowing lateral transshipments between two companies (%SAVE1) is 17.4%. As expected, %SAVE1 is sensitive to the lateral transshipment lead-time and the target service level (maximum waiting time). Intuitively, as lateral transshipment lead-time increases, the savings resulting from the cooperation diminish as more stocks will be needed to satisfy the maximum waiting time constraints. The system with tight waiting time constraints (W^max = 0.1) obtains higher %SAVE1 than
the system with looser waiting time constraints ($W^{\text{max}} = 0.25$). This shows that a lateral transshipment policy becomes more interesting if the cooperating companies set high target service levels.

- The average percentage cost savings when moving from an item approach to a system approach, $\%\text{SAVE}_2$, is 9.15%. It is also shown that the variability of inventory holding costs among items has a significant impact on $\%\text{SAVE}_2$ as indicated in Table 4(c). The percentage cost savings increase when the variability of inventory holding costs is higher. On the other hand, $\%\text{SAVE}_2$ is not sensitive to the variability of ratios between demand and repair rates. This result is in line with the findings of Thonemann et al. (2002).

Table 4. Summary of the computational results

<table>
<thead>
<tr>
<th>(a) Performance measures with respect to $\frac{m_i}{\mu_i}$</th>
<th>(b) Performance measures with respect to $\frac{m_i}{m_2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$U[0.5,2.5]$</td>
<td>$U[0.1,3.0]$</td>
</tr>
<tr>
<td>Total costs</td>
<td>3451500</td>
</tr>
<tr>
<td>$%\text{GAP}$</td>
<td>0.75</td>
</tr>
<tr>
<td>$%\text{SAVE}_1$</td>
<td>17.69</td>
</tr>
<tr>
<td>$%\text{SAVE}_2$</td>
<td>8.70</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(c) Performance measures with respect to $c_i^h$</th>
<th>(d) Performance measures with respect to $\frac{c_i^{m}}{c_i}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$U[5000,15000]$</td>
<td>$U[1000,19000]$</td>
</tr>
<tr>
<td>Total costs</td>
<td>3485900</td>
</tr>
<tr>
<td>$%\text{GAP}$</td>
<td>0.71</td>
</tr>
<tr>
<td>$%\text{SAVE}_1$</td>
<td>17.95</td>
</tr>
<tr>
<td>$%\text{SAVE}_2$</td>
<td>7.50</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(e) Performance measures with respect to $ET_i^r$</th>
<th>(f) Performance measures with respect to $W^{\text{max}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0.1$</td>
<td>$0.25$</td>
</tr>
<tr>
<td>Total costs</td>
<td>3361200</td>
</tr>
<tr>
<td>$%\text{GAP}$</td>
<td>0.79</td>
</tr>
<tr>
<td>$%\text{SAVE}_1$</td>
<td>18.98</td>
</tr>
<tr>
<td>$%\text{SAVE}_2$</td>
<td>8.92</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(g) Performance measures with respect to $N$</th>
<th>(h) Average results for all instances</th>
</tr>
</thead>
<tbody>
<tr>
<td>$20$</td>
<td>$50$</td>
</tr>
<tr>
<td>Total costs</td>
<td>1236100</td>
</tr>
<tr>
<td>$%\text{SAVE}_1$</td>
<td>0.99</td>
</tr>
<tr>
<td>$%\text{SAVE}_2$</td>
<td>17.60</td>
</tr>
</tbody>
</table>
• The total costs values increase linearly with the number of items. The ratios between emergency supply costs and lateral transshipment costs, the lateral transshipment lead times, and the maximum waiting times are factors that also influence the total costs. The other parameters, the variability of the inventory holding costs, the variability of the ratios between demand rates and repair rates, and also the relative sizes of the two companies have only a very small effect on total costs.

The average computation times for solving the problems in Table 4 are 2.5, 6.3 and 13.8 minutes for \( N = 20, 50 \) and 100 respectively, using a PC with a 333-MHz Pentium II processor.

5. Model application

As already mentioned, this research was originally motivated by an air carrier company located in Brussels who wanted to cooperate with another company to pool their spare parts inventories. We performed a pilot study to help management of the company to have an idea of the cost advantages of a pooling strategy. A potential partner considered at that time was a company located in Liege which has approximately two hours of driving from Brussels. However, since complete information of the partner company was not available, we assumed that both companies were identical. We selected a sample of 32 expensive parts, of which the prices are at least \( \geq 25,000 \). The maximum average waiting time, as desired by the company, was set at 2 hours. Holding costs were 20% of unit prices. Transportation costs per unit for lateral transshipments are based on the distance between the two companies. Three possible distances: 2, 4, and 6 hours were selected (the two latter distances are used for the purpose of sensitivity analysis). The unit transportation cost was set at \( \varepsilon 50 \) per hour. Emergency supply lead-time was set at the average level of one day with the costs of \( \varepsilon 500 \). Table 4 shows all the data and solutions for this model application. In particular, we compared the total costs resulting from the pooling policy with the total costs corresponding to the current company’s no-cooperation policy. For the range of distances between 2 and 6 hours, the gained savings range from 21% to 14.5%. In conclusion, the use of a pooling strategy can reduce their respective total annual operating costs significantly. The percentage gaps between the lower and upper bounds for the three distances are 1.19%, 1.05%, and 1.03%.
Table 4. Data and solutions for model application

<table>
<thead>
<tr>
<th>Part #</th>
<th>Part title</th>
<th>$m_i$ (day$^{-1}$)</th>
<th>$\mu_i$ (day$^{-1}$)</th>
<th>Price (€)</th>
<th>No pooling</th>
<th>$ET_i^{m=2}$</th>
<th>$ET_i^{m=4}$</th>
<th>$ET_i^{m=6}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>flap elec. control</td>
<td>0.0229</td>
<td>0.0141</td>
<td>143450</td>
<td>2, 2</td>
<td>1, 1</td>
<td>1, 1</td>
<td>2, 1</td>
</tr>
<tr>
<td>2</td>
<td>flap hydr. control</td>
<td>0.0029</td>
<td>0.0204</td>
<td>113625</td>
<td>0, 0</td>
<td>0, 0</td>
<td>0, 0</td>
<td>0, 0</td>
</tr>
<tr>
<td>3</td>
<td>fuel control</td>
<td>0.0143</td>
<td>0.0263</td>
<td>94575</td>
<td>1, 1</td>
<td>1, 1</td>
<td>1, 1</td>
<td>1, 1</td>
</tr>
<tr>
<td>4</td>
<td>autopilot computer</td>
<td>0.0343</td>
<td>0.0303</td>
<td>89900</td>
<td>3, 3</td>
<td>2, 2</td>
<td>3, 2</td>
<td>3, 2</td>
</tr>
<tr>
<td>5</td>
<td>engine display</td>
<td>0.0114</td>
<td>0.0417</td>
<td>74875</td>
<td>1, 1</td>
<td>1, 1</td>
<td>1, 1</td>
<td>1, 1</td>
</tr>
<tr>
<td>6</td>
<td>APU generator gearbox</td>
<td>0.0029</td>
<td>0.0084</td>
<td>55750</td>
<td>0, 0</td>
<td>1, 0</td>
<td>1, 0</td>
<td>0, 0</td>
</tr>
<tr>
<td>7</td>
<td>cold air unit</td>
<td>0.0029</td>
<td>0.0119</td>
<td>50292</td>
<td>1, 1</td>
<td>1, 0</td>
<td>1, 0</td>
<td>1, 0</td>
</tr>
<tr>
<td>8</td>
<td>multi process unit</td>
<td>0.0400</td>
<td>0.0500</td>
<td>47775</td>
<td>3, 3</td>
<td>3, 2</td>
<td>3, 2</td>
<td>3, 2</td>
</tr>
<tr>
<td>9</td>
<td>actuator flap nacelle</td>
<td>0.0029</td>
<td>0.0217</td>
<td>45750</td>
<td>1, 1</td>
<td>1, 0</td>
<td>1, 0</td>
<td>1, 0</td>
</tr>
<tr>
<td>10</td>
<td>pitch computer</td>
<td>0.0171</td>
<td>0.0294</td>
<td>45325</td>
<td>2, 2</td>
<td>2, 1</td>
<td>2, 2</td>
<td>2, 2</td>
</tr>
<tr>
<td>11</td>
<td>radar</td>
<td>0.0086</td>
<td>0.0345</td>
<td>44100</td>
<td>1, 1</td>
<td>1, 1</td>
<td>1, 1</td>
<td>1, 1</td>
</tr>
<tr>
<td>12</td>
<td>roll computer</td>
<td>0.0057</td>
<td>0.0185</td>
<td>43350</td>
<td>1, 1</td>
<td>1, 1</td>
<td>1, 1</td>
<td>1, 1</td>
</tr>
<tr>
<td>13</td>
<td>rudder servo</td>
<td>0.0057</td>
<td>0.0167</td>
<td>41325</td>
<td>1, 1</td>
<td>1, 1</td>
<td>1, 1</td>
<td>1, 1</td>
</tr>
<tr>
<td>14</td>
<td>TMS computer</td>
<td>0.0400</td>
<td>0.0147</td>
<td>40225</td>
<td>6, 6</td>
<td>5, 4</td>
<td>5, 5</td>
<td>5, 5</td>
</tr>
<tr>
<td>15</td>
<td>generator</td>
<td>0.0457</td>
<td>0.0164</td>
<td>37550</td>
<td>6, 6</td>
<td>5, 5</td>
<td>5, 5</td>
<td>5, 5</td>
</tr>
<tr>
<td>16</td>
<td>flap actuator outboard</td>
<td>0.0457</td>
<td>0.0357</td>
<td>34800</td>
<td>4, 4</td>
<td>3, 3</td>
<td>4, 3</td>
<td>4, 3</td>
</tr>
<tr>
<td>17</td>
<td>roll splr. servo</td>
<td>0.0114</td>
<td>0.0139</td>
<td>33925</td>
<td>2, 2</td>
<td>2, 2</td>
<td>2, 2</td>
<td>2, 2</td>
</tr>
<tr>
<td>18</td>
<td>cabin pressure control</td>
<td>0.0086</td>
<td>0.0159</td>
<td>32475</td>
<td>2, 2</td>
<td>2, 1</td>
<td>2, 1</td>
<td>2, 1</td>
</tr>
<tr>
<td>19</td>
<td>instr. switching unit</td>
<td>0.0257</td>
<td>0.0500</td>
<td>31425</td>
<td>3, 3</td>
<td>2, 2</td>
<td>2, 2</td>
<td>2, 2</td>
</tr>
<tr>
<td>20</td>
<td>prop. control unit</td>
<td>0.0171</td>
<td>0.0125</td>
<td>30750</td>
<td>3, 3</td>
<td>3, 3</td>
<td>3, 3</td>
<td>3, 3</td>
</tr>
<tr>
<td>21</td>
<td>distance bearing</td>
<td>0.0143</td>
<td>0.0130</td>
<td>30025</td>
<td>3, 3</td>
<td>3, 2</td>
<td>3, 2</td>
<td>3, 2</td>
</tr>
<tr>
<td>22</td>
<td>servo altimeter</td>
<td>0.0400</td>
<td>0.0278</td>
<td>29000</td>
<td>4, 4</td>
<td>4, 3</td>
<td>4, 3</td>
<td>4, 4</td>
</tr>
<tr>
<td>23</td>
<td>speed brake hydr. act.</td>
<td>0.0086</td>
<td>0.0417</td>
<td>28700</td>
<td>2, 2</td>
<td>1, 1</td>
<td>1, 1</td>
<td>1, 1</td>
</tr>
<tr>
<td>24</td>
<td>navigation selector</td>
<td>0.0171</td>
<td>0.0370</td>
<td>28325</td>
<td>2, 2</td>
<td>2, 2</td>
<td>2, 2</td>
<td>2, 2</td>
</tr>
<tr>
<td>25</td>
<td>yaw actuator</td>
<td>0.0114</td>
<td>0.0167</td>
<td>27625</td>
<td>2, 2</td>
<td>2, 2</td>
<td>2, 2</td>
<td>2, 2</td>
</tr>
<tr>
<td>26</td>
<td>display processor unit</td>
<td>0.0886</td>
<td>0.0476</td>
<td>27425</td>
<td>6, 6</td>
<td>5, 4</td>
<td>5, 5</td>
<td>5, 5</td>
</tr>
<tr>
<td>27</td>
<td>flap indicator switch</td>
<td>0.0029</td>
<td>0.0213</td>
<td>27375</td>
<td>1, 1</td>
<td>1, 0</td>
<td>1, 0</td>
<td>1, 1</td>
</tr>
<tr>
<td>28</td>
<td>flap screw jack</td>
<td>0.0171</td>
<td>0.0500</td>
<td>27050</td>
<td>2, 2</td>
<td>2, 1</td>
<td>2, 1</td>
<td>2, 2</td>
</tr>
<tr>
<td>29</td>
<td>EFIS display</td>
<td>0.0286</td>
<td>0.0196</td>
<td>26075</td>
<td>4, 4</td>
<td>4, 3</td>
<td>4, 3</td>
<td>4, 3</td>
</tr>
<tr>
<td>30</td>
<td>TMS display</td>
<td>0.0200</td>
<td>0.0167</td>
<td>25650</td>
<td>4, 4</td>
<td>3, 3</td>
<td>3, 3</td>
<td>3, 3</td>
</tr>
<tr>
<td>31</td>
<td>discharge valve</td>
<td>0.0086</td>
<td>0.0137</td>
<td>25250</td>
<td>2, 2</td>
<td>2, 1</td>
<td>2, 2</td>
<td>2, 2</td>
</tr>
<tr>
<td>32</td>
<td>air speed indicator</td>
<td>0.0143</td>
<td>0.0192</td>
<td>25000</td>
<td>3, 3</td>
<td>2, 2</td>
<td>2, 2</td>
<td>2, 2</td>
</tr>
</tbody>
</table>

Total costs: 1244700, 973880, 1028100, 1064700

%Savings: 21.97%, 17.04%, 14.46%

%GAP: 1.19%, 1.05%, 1.03%
6. Conclusions and directions for further research

In this article we considered a multi-item, continuous review model of a two-location inventory system for repairable spare parts in which lateral and emergency shipments can occur in response to stockouts. In the system that we analyzed the failed part at one location is replaced by a ready-for-use spare part that can come either from the local warehouse, or from the other location as a lateral transshipment, or from an emergency supply. Additionally, our formulation addresses the concern for service performance by stating constraints in terms of maximum average waiting time for a ready-for-use spare part. We have developed a solution procedure based on Lagrangian relaxation that provides tight bounds of the optimal total costs.

Our computational results give us some important insights into the tightness of the bounds and into the effect of the system parameters on the system performance that might be of interest from a managerial point of view. Some of which are summarized as follows: (1) the quality of our heuristic solution is quite good as indicated by a very small percentage of the gap between the costs of our solution and the lower bound for the optimal costs, (2) the performance of our solution increases with an increasing number of items, (3) the relative cost savings of the pooling policy increase when the lateral transshipment lead-time is short and the target service level is high, and (4) the relative cost savings of using a system approach increase when the variability of inventory holding costs among items increases.

The application of our model in the real system of an air carrier company has indicated that significant cost savings can be gained by pooling the spare parts inventories with another company.

Our work can be extended in several directions. One possible extension is to consider more than two, say $N$ companies. The main difficulty is defining a policy for lateral transshipments when there may be more than one company that can be the source for the lateral transshipments. A very reasonable policy is to source the lateral transshipment from the closest neighbor company. In fact, under such pre-specified policy one can apply the same optimization procedure as in this paper. However, as the number of companies grows, the development of a more efficient heuristic is needed and becomes an interesting topic for further research.
References


Tagaras, G., 1989. Effects of pooling on the optimization and service levels of two-location inventory systems. IIE Transactions 21, 250-257.


