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

Multi-Item Spare Parts Systems with Lateral Transshipments and Waiting Time Constraints

H. Wong [a*], G. J. van Houtum [b], D. Cattrysse [a] and D. Van Oudheusden [a]

[a] *Katholieke Universiteit Leuven, Centre for Industrial Management, Celestijnenlaan 300A, 3001 Heverlee, Belgium*
[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

* Corresponding author. Tel.: +32-16-322497; fax: +32-16-322986
E-mail address: hartanto.wong@cib.kuleuven.ac.be

down. 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), Axsäter (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 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 1. A brief review of the literature on the multi-location inventory systems

paper	# echelons	# items	class of policies	model analysis	important remarks
Gross (1963)	single	single	(s, S)	exact opt.	• allows the transshipments before the realization of demand
Krishnan and Rao (1965)	single	single	(s, S)	exact. opt.	• derive optimal solution for two-location systems with identical cost parameters; allow the transshipment after demand is realized but before it must be satisfied
Das (1975)	single	single	(s, S)	exact opt.	• extends Gross's model by permitting transshipments in the middle of each period
Hoadley and Heyman (1977)	two	single	(s, S)	exact opt.	• allow balancing acts at the beginning of a period and emergency acts at the end of a period
Karmarkar (1981)	single	single	general	appr. opt.	• considers the multi-period problem; develops lower and upper bounds of the optimal costs
Cohen et al. (1986)	multi	single	(s, S)	appr. eval.; appr. opt.	• allow transshipments at higher echelons
Lee (1987)	two	single	(S-1, S)	appr. eval.; appr. opt.	• assumes identical bases
Tagaras (1989)	single	single	(s, S)	appr. opt.	• extends the model of Krishnan and Rao (1965) by using a more general cost structure
Axsäter (1990)	two	single	(S-1, S)	appr. eval.	• allows non-identical bases, focus on the demand processes
Robinson (1990)	single	single	general	exact opt.	• 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
Tagaras and Cohen (1992)	single	single	(s, S)	appr. eval.; appr. opt.	• consider positive replenishment lead times and shows the superiority of complete pooling
Yanagi and Sasaki (1992)	single	single	(S-1, S)	appr. eval.	• consider limited repair capacity
Sherbrooke (1992)	two	single	(S-1, S)	appr. eval.	• estimates the expected backorders using a regression model
Archibald et al. (1997)	single	multi	(s, S)	exact. eval; exact opt.	• allow lateral transshipments at any time during a period; model the problem as a Markov decision process • considers the multi-item problem with limited storage space
Needham and Evers (1998)	two	single	(R, Q)	appr. eval.; appr. opt.	• examine the interaction of relevant costs and transshipment policies; use a simulation-based optimization
Tagaras (1999)	single	single	(s, S)	appr. eval.	• studies several lateral transshipment policies using a simulation model
Herer and Rashit (1999)	single	single	general	exact opt.	• consider fixed and joint replenishment costs; characterize the form of the optimal policy (single period)
Alfredsson and Verrijdt (1999)	two	single	(S-1, S)	appr. eval.	• allow emergency shipments from the depot and an outside supplier
Grahovac & Chakravarty (2001)	two	single	(S-1, S)	appr. eval.	• allow lateral transshipments before and after a location is out of stock
Kukreja et al. (2001)	single	single	(S-1, S)	appr. eval.; appr. opt.	• relax the assumption of exponential repair time distribution
Evers (2001)	single	single	(R, Q)	appr. opt.	• develops heuristics for determining when transshipments should be made
Rudi et al. (2001)	single	single	(s, S)	exact. opt.	• consider local decision making; focus on determining transshipment price
Wong et al. (2002)	single	single	(S-1, S)	appr. eval.; appr. opt.	• consider pooling in a multi-hub system and delayed lateral transshipments
Herer et al. (2002)	single	single	general	appr. opt.	• prove the optimality of (s,S) policies; consider dynamic transshipments; use a simulation-based optimization
Xu et al. (2003)	single	single	(R, Q)	appr. eval.	• introduce hold-back parameter which limit the level of outgoing transshipments
Axsäter (2003)	single	single	(R, Q)	appr. eval.	• allow lateral transshipments only in one direction
This paper	single	multi	(S-1, S)	exact eval.; appr.opt.	• considers the multi-item problem and employs waiting time constraints

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} (≥ 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 $m_{i1} + m_{i2} > 0$ 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 $S_{ij} (\in \mathbb{N}_0 := \{0\} \cup \mathbb{N})$ spare parts of item i. We define $\underline{S}_j := (S_{1j}, \ldots, S_{Ij})$. In total, both companies share S_i spare parts of item i, where $S_i = S_{i1} + S_{i2}$. 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_i^{tr} (> 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 $ET_i^{em} (\geq ET_i^{tr})$. 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 μ_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^{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_i^h are counted for each spare part of item i. A cost c_i^{tr} is counted for each lateral transshipment of part of item i. A cost c_i^{em} is counted for each part coming from the emergency supply. The objective is to find a policy $(\underline{S}_1, \underline{S}_2)$ under which the total average costs are minimized subject to the waiting time constraints for the companies 1 and 2.

2.2. Model formulation

To formulate the problem, we define:

β_{ij} = fraction of demands for item i at company j satisfied by its own stocks

α_{ij} = fraction of demands for item i at company j satisfied by lateral transshipments

θ_{ij} = fraction of demands for item i at company j satisfied by emergency supply

W_j = 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, θ_{ij} is the same for $j = 1, 2$, i.e. $\theta_{i1} = \theta_{i2} = \theta_i$ 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 $\boldsymbol{x_i} = (x_{i1},x_{i2})$, where x_{ij} represents the physical stock of spare parts of item i at company j, and $0 \le x_{ij} \le S_{ij}$, $x_{ij} \in \mathbb{N}_0$. We define $\boldsymbol{x_{i1-}} = (x_{i1}-1,x_{i2})$, $\boldsymbol{x_{i1+}} = (x_{i1}+1,x_{i2})$, $\boldsymbol{x_{i2-}} = (x_{i1},x_{i2}-1)$, $\boldsymbol{x_{i2+}} = (x_{i1},x_{i2}+1)$. 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 $\boldsymbol{x_i} \rightarrow \boldsymbol{x_{ij-}}$; 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 $\boldsymbol{x_i} \rightarrow \boldsymbol{x_{ij'-}}$; 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 $x_{i1}=x_{i2}=0$; an emergency supply is applied; the state transition is $\boldsymbol{x_i} \rightarrow \boldsymbol{x_i}$; 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 $\boldsymbol{x_i} \rightarrow \boldsymbol{x_{ij+}}$; the transition rate is $(S_{ij}-x_{ij})\mu_i$.

Figure 1 shows the Markov process that is obtained when $(S_{i1},S_{i2}) = (2,1)$. A similar process is obtained for any other choice for (S_{i1},S_{i2}).

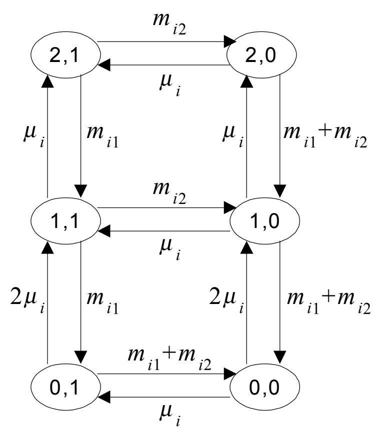

Figure 1. Markov process for item i with $(S_{i1},S_{i2}) = (2,1)$

We define $\boldsymbol{\pi}$ as the steady-state probability vector and $\pi_{(k,l)}$ as the steady-state probability of being in state (k,l), $0 \le k \le S_{i1}$, $0 \le l \le S_{i2}$. Since the number of states in our problem is not large, a direct method based on Gaussian elimination is applied to determine $\boldsymbol{\pi}$. 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 θ_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 $\alpha_{i1} = \sum_{l=1}^{S_{i2}} \pi_{(0,l)}$. Similarly, $\alpha_{i2} = \sum_{k=1}^{S_{i1}} \pi_{(k,0)}$. The fraction of demands that is satisfied by the local stock is obtained from $\beta_{ij} = 1 - \alpha_{ij} - \theta_i$, j = 1,2.

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:

$W_j = \sum_{i=1}^{I}$ *Prob*{an arbitrary failing part at location *j* is of item *i*} (average waiting time for a ready-for-use part of item *i*)

$$= \sum_{i=1}^{I} \frac{m_{ij}}{M_j} (\beta_{ij} 0 + \alpha_{ij} ET_i^{tr} + \theta_i ET_i^{em})$$

$$= \sum_{i=1}^{I} \frac{m_{ij}}{M_j} (\alpha_{ij} ET_i^{tr} + \theta_i ET_i^{em}) \tag{1}$$

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

Problem P_0: Minimize $\sum_{j=1}^{2} \sum_{i=1}^{I} \left(c_i^h S_{ij} + c_i^{tr} m_{ij} \alpha_{ij} + c_i^{em} m_{ij} \theta_i \right)$ (2)

subject to $\sum_{i=1}^{I} \frac{m_{ij}}{M_j} (\alpha_{ij} ET_i^{tr} + \theta_i ET_i^{em}) \le W_j^{max}$, $j = 1, 2$, (3)

$S_{ij} \in \mathbb{N}_0$, $i = 1, \ldots, I$; $j = 1, 2$ (4)

The constraints in (3) can be rewritten as

$$\sum_{i=1}^{I} m_{ij} (\alpha_{ij} ET_i^{tr} + \theta_i ET_i^{em}) \le M_j W_j^{max} \quad j = 1, 2 \tag{5}$$

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 $\lambda \in \Re^2$ 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

$$\sum_{j=1}^{2}\sum_{i=1}^{I}\left(c_i^h S_{ij} + c_i^{tr} m_{ij}\alpha_{ij} + c_i^{em} m_{ij}\theta_i\right) + \sum_{j=1}^{2}\lambda_j\left(\sum_{i=1}^{I} m_{ij}(\alpha_{ij}ET_i^{tr} + \theta_i ET_i^{em}) - M_j W_j^{max}\right) \quad (6)$$

subject to $\quad S_{ij} \in \mathbb{N}_0 \quad i = 1, \ldots, I;\ j = 1, 2 \qquad (7)$

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^{*P_0} denote the costs of the optimal solution of problem P_0, let $Z^{P_1}(\lambda)$ denote the costs of the optimal solution of problem P_1 for given $\lambda = (\lambda_1,\lambda_2)$, and let $Z^{P_0}(\underline{S}_1,\underline{S}_2)$ denote the costs of problem P_0 under basestock policy $(\underline{S}_1,\underline{S}_2)$. We define $W_j(\underline{S}_1,\underline{S}_2)$ as the average waiting time for company j obtained under the stocking policy $(\underline{S}_1,\underline{S}_2)$.

Property 1

(i) $\quad Z^{*P_0} \geq Z^{P_1}(\lambda)$ for all $(\lambda_1,\lambda_2) \geq (0,0)$

(ii) $\quad Z^{*P_0} \geq Max_\lambda\ Z^{P_1}(\lambda)$

(iii) If for some $\lambda = (\lambda_1,\lambda_2) \geq (0,0)$ the optimal solution for problem P_1 is $(\underline{S}_1^*,\underline{S}_2^*)$ and $W_j(\underline{S}_1^*,\underline{S}_2^*) \leq W_j^{max}$, j = 1, 2, then $(\underline{S}_1^*,\underline{S}_2^*)$ is feasible for problem P_0, and $Z^{P_0}(\underline{S}_1^*,\underline{S}_2^*) - Z^{*P_0} \leq \lambda_1(W_1^{max} - W_1(\underline{S}_1^*,\underline{S}_2^*)) + \lambda_2(W_2^{max} - W_2(\underline{S}_1^*,\underline{S}_2^*))$.

(iv) If for some $\lambda = (\lambda_1,\lambda_2) \geq (0,0)$ the optimal solution for problem P_1 is $(\underline{S}_1^*,\underline{S}_2^*)$ and $W_j(\underline{S}_1^*,\underline{S}_2^*) = W_j^{max}$, j = 1, 2, then $(\underline{S}_1^*,\underline{S}_2^*)$ 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 (λ_1,λ_2) we obtain a lower bound on Z^{*P_0}. The maximum value of this lower bound over all considered values (λ_1,λ_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 λ

The objective function of P_1 may be rewritten as

$$\text{Minimize} \quad \sum_{i=1}^{I}\sum_{j=1}^{2}\left(c_i^h S_{ij} + (c_i^{tr} + \lambda_j ET_i^{tr})m_{ij}\alpha_{ij} + (c_i^{em} + \lambda_j ET_i^{em})m_{ij}\theta_i\right) - \sum_{j=1}^{2}\lambda_j M_j W_j^{max} \quad (8)$$

Now, the λ_j values are assumed to be given. Then the second factor $-\sum_{j=1}^{2}\lambda_j M_j W_j^{max}$ 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 $\quad c_i^h S_{i1} + (c_i^{tr} + \lambda_1 ET_i^{tr}) m_{i1}\alpha_{i1} + (c_i^{em} + \lambda_1 ET_i^{em}) m_{i1}\theta_i$

$$+c_i^h S_{i2} + (c_i^{tr} + \lambda_2 ET_i^{tr}) m_{i2}\alpha_{i2} + (c_i^{em} + \lambda_2 ET_i^{em}) m_{i2}\theta_i \qquad (9)$$

subject to $\quad S_{i1} \in \mathbb{N}_0,\ S_{i2} \in \mathbb{N}_0$

Recall that α_{i1} and α_{i2} depend on the individual stocks S_{i1} and S_{i2}, while θ_i depends on the aggregate stocks level S_i . We can rewrite the objective function as follows:

Minimize $\quad f(S_i) + g(S_{i1},S_{i2}) \qquad (10)$

where $\quad f(S_i) = c_i^h S_i + (c_i^{em} m_{i1} + \lambda_1 m_{i1} ET_i^{em} + c_i^{em} m_{i2} + \lambda_2 m_{i2} ET_i^{em})\theta_i \qquad (11)$

$$g(S_{i1},S_{i2}) = (c_i^{tr} + \lambda_1 ET_i^{tr}) m_{i1}\alpha_{i1} + (c_i^{tr} + \lambda_2 ET_i^{tr}) m_{i2}\alpha_{i2}\,. \qquad (12)$$

First, we will look at the behavior of the costs function. It is known that θ_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_{i1},S_{i2}) with $S_{i1} + S_{i2} = S_i$ such that $g(S_{i1},S_{i2})$ is minimized. Let us define $g^*(S_i) = \underset{S_{i1},S_{i2}}{\text{Min}}\{g(S_{i1},S_{i2}) \mid S_{i1} + S_{i2} = S_i\}$. 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 $S_i = 0$, $S_{i1} = 0$, $S_{i2} = 0$, min $= f(0)$, $S_{i1}^* = S_{i1}$, $S_{i2}^* = S_{i2}$.

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 $S_{i1} = k$ and $S_{i2} = S_i - k$; calculate $f(S_i)$ and $g(S_{i1},S_{i2})$.

Step 4: If $f(S_i) + g(S_{i1}, S_{i2}) <$ min, then set min $= f(S_i) + g(S_{i1}, S_{i2})$, $S_{i1}^* = S_{i1}$ and $S_{i2}^* = S_{i2}$.

Step 5: If $k < S_i$, set $k = k + 1$ and go to *Step 3*, otherwise go to *Step 2*.

END

3.3. Finding the tightest lower bound

Once the Lagrangian problem P_1 is solved for a given λ, we know that the objective function $Z^{P_1}(\lambda)$ 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 $Z^{P_1}(\lambda^*)$ where

$$Z^{P_1}(\lambda^*) = \underset{\lambda \geq 0}{Max}\, Z^{P_1}(\lambda). \tag{13}$$

The next step is to find the best Lagrange multipliers λ_1^* and λ_2^*. Since $Z^{P_1}(\lambda)$ 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. Bazaraa 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 λ^k at iteration k of the subgradient method. Then the subgradient direction $\gamma^k = (\gamma_1^k, \gamma_2^k)$ for λ^k is calculated as

$$\gamma_j^k = W_j^k - W_j^{max} \quad \text{for } j = 1, 2. \tag{14}$$

At each iteration k, the vector of Lagrange multipliers for iteration $(k + 1)$ is updated as

$$\lambda_j^{k+1} = \max\left(0, \lambda_j^k - t^k \gamma_j^k\right) \quad \text{for } j = 1, 2. \tag{15}$$

In this formula, t^k is a scalar stepsize. For the t^k, we use

$$t^k = s^k \frac{Z^{P_1}(\lambda^k) - \widehat{Z}}{\left\|\gamma^k\right\|_2^2} \tag{16}$$

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, $\widehat{Z}$ 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 $\lambda^0 = \mathbf{0}$ 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 λ^*. We describe the initialization procedure in the following part.

Initialization procedure

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

Property 2:

Let $(\underline{S}_1^*, \underline{S}_2^*)_{\lambda_1,\lambda_2}$ denote the optimal solution of problem P_1 at penalty values λ_1 and λ_2, and $W_j(\underline{S}_1^*, \underline{S}_2^*)_{\lambda_1,\lambda_2}$, j = 1, 2, the corresponding achieved average waiting times.

(i) For $\lambda_1^{'} > \lambda_1$, $W_1(\underline{S}_1^*, \underline{S}_2^*)_{\lambda_1^{'},\lambda_2} \le W_1(\underline{S}_1^*, \underline{S}_2^*)_{\lambda_1,\lambda_2}$

(ii) For $\lambda_2^{'} > \lambda_2$, $W_2(\underline{S}_1^*, \underline{S}_2^*)_{\lambda_1,\lambda_2^{'}} \le W_2(\underline{S}_1^*, \underline{S}_2^*)_{\lambda_1,\lambda_2}$

Proof: See Theorem 2, Van Houtum and Zijm (2002) .

For each value of λ_2 there exists a smallest value of λ_1, say $\lambda_{1\min}$, at which W_1^{max} is satisfied. These points form a function in λ_2. Similarly, we can have $\lambda_{2\min}$ as a function of λ_1. From Property 2, it follows that of all the points (λ_1,λ_2) that give feasible solutions for problem P_0, the point where the functions $\lambda_{1\min}(\lambda_2)$ and $\lambda_{2\min}(\lambda_1)$ cross each other (if they do cross) is the best estimate for λ^*. If both functions do not cross, we would expect that the location of λ^* is close to the point (λ_1,λ_2) where both functions come close to each other. Let us define $(\hat{\lambda}_{1\min},\hat{\lambda}_{2\min})$ to represent such point.

We performed a computational experiment to learn about the behavior of the functions $\lambda_{1\min}(\lambda_2)$, $\lambda_{2\min}(\lambda_1)$, and the locations of $(\hat{\lambda}_{1\min},\hat{\lambda}_{2\min})$. 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 $\lambda_{1\min}(\lambda_2)$ and $\lambda_{2\min}(\lambda_1)$ and also $(\hat{\lambda}_{1\min},\hat{\lambda}_{2\min})$. In each figure, $(\hat{\lambda}_{1\min},\hat{\lambda}_{2\min})$ is represented by a small circle. As expected, we found that in general, $\lambda_{1\min}$ is decreasing in λ_2, and $\lambda_{2\min}$ is decreasing in λ_1. The locations of $(\hat{\lambda}_{1\min},\hat{\lambda}_{2\min})$ can be classified as follows:

(a) $\hat{\lambda}_{1\min} \approx \hat{\lambda}_{2\min}$ (see Figure 2(a) and 2(b)): this occurs when the two companies are identical so that the function $\lambda_{1\min}(\lambda_2)$ and $\lambda_{2\min}(\lambda_1)$ 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 $\lambda_{1\min}(\lambda_2)$ and $\lambda_{2\min}(\lambda_1)$ lie on the same line as shown by Figure 1(a).

(b) $\hat{\lambda}_{1\min} < \hat{\lambda}_{2\min}$ (see Figure 2(c) and 2(d)): this occurs in the two following cases: (i) $m_{i1} > m_{i2}$ for all i and $W_1^{max} = W_2^{max}$; (ii) $m_{i1} \geq m_{i2}$ for all i and $W_1^{max} > W_2^{max}$. In those two cases, if ℓ is as an arbitrary value of Lagrange multiplier, we will find that $\lambda_{1\min}(\lambda_2{=}\ell) < \lambda_{2\min}(\lambda_1{=}\ell)$. In the extreme case, we may find the situation where the first constraint is inactive and $\hat{\lambda}_{1\min} = 0$.

(c) $\hat{\lambda}_{1\min} > \hat{\lambda}_{2\min}$ (see Figure 2(e) and 2(f)): this occurs in the two following cases: (i) $m_{i1} < m_{i2}$ and $W_1^{max} = W_2^{max}$; (ii) $m_{i1} \leq m_{i2}$ and $W_1^{max} < W_2^{max}$. 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 (λ_1,λ_2). For the first point, λ_1 is set to zero, and we determine smallest value λ_2 for which both constraints are satisfied. For the second point, we have to find the smallest value λ, where $\lambda_1 = \lambda_2 = \lambda$, such that both constraints are satisfied. And for the third point, similar to the first point, we set λ_2 equal to zero and we determine smallest value λ_1 for which both constraints are satisfied. The point with the largest objective function value $Z^{P_1}(\lambda)$ 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^{P_0} of this solution is used as the initial value for $\hat{Z}$ in the subgradient method.

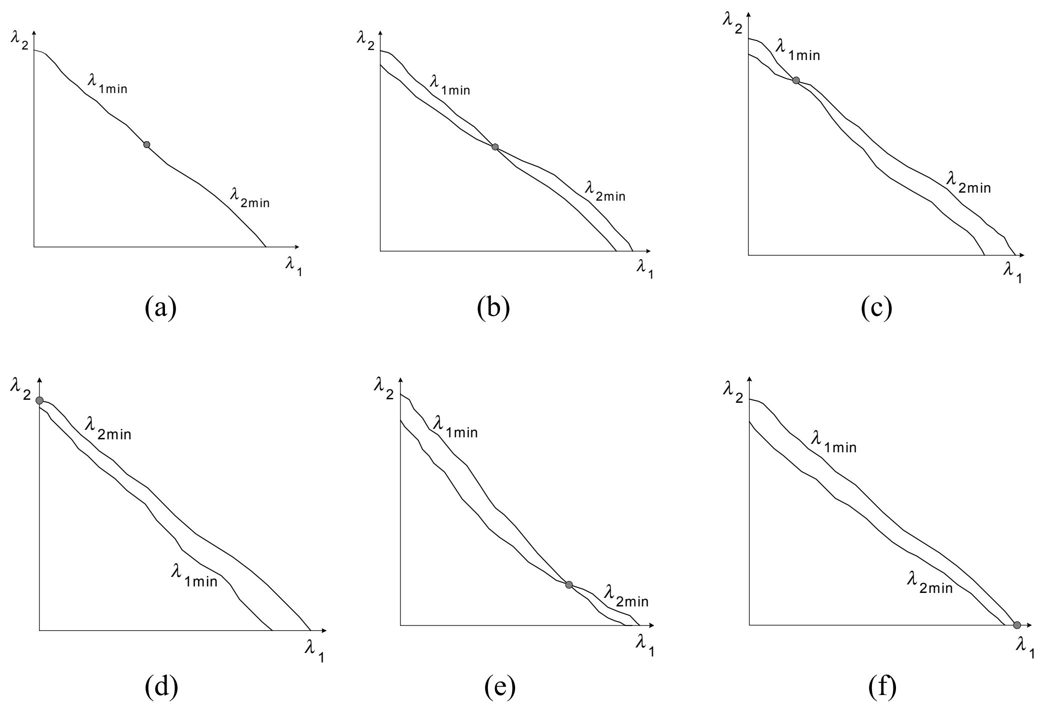

Figure 2. Some possibilities of $\lambda_{1\min}(\lambda_2)$, $\lambda_{2\min}(\lambda_1)$ and $(\hat{\lambda}_{1\min}, \hat{\lambda}_{2\min})$

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 $(\underline{S}_1, \underline{S}_2)$ that is feasible in problem P_0, we evaluate the costs $Z^{P_0}(\underline{S}_1, \underline{S}_2)$ 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 $(\underline{S}_1, \underline{S}_2)$ with $W_1(\underline{S}_1, \underline{S}_2) > W_1^{\max}$ or $W_2(\underline{S}_1, \underline{S}_2) > W_2^{\max}$.

Step 1: If $W_1(\underline{S}_1, \underline{S}_2) > W_1^{\max}$ and $W_2(\underline{S}_1, \underline{S}_2) > W_2^{\max}$ go to *Step 4*, otherwise continue.

Step 2: Here we have either (*a*) $W_1(\underline{S}_1, \underline{S}_2) > W_1^{\max}$ and $W_2(\underline{S}_1, \underline{S}_2) \le W_2^{\max}$; or (*b*) $W_1(\underline{S}_1, \underline{S}_2) \le W_1^{\max}$ and $W_2(\underline{S}_1, \underline{S}_2) > W_2^{\max}$. For (*a*), find item *k* such that $W_2^{\max} - W_{k2}(S_{k1}, S_{k2}) = \max(W_2^{\max} - W_{12}(S_{11}, S_{12}), ..., W_2^{\max} - W_{I2}(S_{I1}, S_{I2}))$. Set $S_{k2} = S_{k2} - 1$ and $S_{k1} = S_{k1} + 1$. For (*b*), do the symmetry.

Step 3: Calculate $W_{k1}(S_{k1}, S_{k2})$, $W_{k2}(S_{k1}, S_{k2})$, $W_1(\underline{S}_1, \underline{S}_2)$ and $W_2(\underline{S}_1, \underline{S}_2)$. If $W_1(\underline{S}_1, \underline{S}_2) > W_1^{\max}$ and $W_2(\underline{S}_1, \underline{S}_2) > W_2^{\max}$ go to *Step 4*, otherwise if $W_1(\underline{S}_1, \underline{S}_2) \le W_1^{\max}$ and $W_2(\underline{S}_1, \underline{S}_2) \le W_2^{\max}$ go to *END*, otherwise go *to Step 2*.

Step 4: For each item *i*, obtain $r^i = \dfrac{(m_{i1} + m_{i2}) ET_i^{em} \left(\theta_i(S_i) - \theta_i(S_i + 1)\right)}{c_i^h}$ and choose item *k* with $r^k = \max(r^1, ..., r^I)$. If $W_1^{\max} - W_1(\underline{S}_1, \underline{S}_2) < W_2^{\max} - W_2(\underline{S}_1, \underline{S}_2)$ 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 μ_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, μ_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^h, ten samples were generated. This results in 120 sample sets. Combined with 2x2x2x2=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:

$$\%GAP = \frac{\text{upper bound - lower bound}}{\text{lower bound}} \text{ x } 100$$

- *%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

Name of the parameter	Unit	Number of values	Values
Number of items (N)		3	20, 50, 100
Inventory holding costs (c_i^h)	\$/unit/year	2	U[5000,15000], U[1000,19000]
Transshipment costs (c_i^{tr})	\$	1	250
Emergency supply costs (c_i^{em})	\$	2	1250, 2500
Lateral transshipment lead time (ET_i^{tr})	days	2	0.1, 0.25
Emergency supply lead time (ET_i^{em})	days	1	2
Maximum waiting time ($W_1^{max} = W_2^{max}$)	days	2	0.25, 0.1
$\frac{\text{Demand rates}}{\text{Repair rates}}$ ($\frac{m_{i1}+m_{i2}}{\mu_i}$)		2	U[0.5,2.5], U[0.1,3.0]
$\frac{\text{Demand rates at company 1}}{\text{Demand rates at company2}}$ ($\frac{m_{i1}}{m_{i2}}$)		2	1, 3

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^{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, *%SAVE2* is 9.15%. It is also shown that the variability of inventory holding costs among items has a significant impact on *%SAVE2* as indicated in Table 4(c). The percentage cost savings increase when the variability of inventory holding costs is higher. On the other hand, *%SAVE2* 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

(a) Performance measures with respect to $\frac{m_i}{\mu_i}$

	U[0.5,2.5]	U[0.1,3.0]
Total costs	3451500	3472500
%GAP	0.75	0.78
%SAVE1	17.69	17.17
%SAVE2	8.70	9.60

(b) Performance measures with respect to $\frac{m_{i1}}{m_{i2}}$

	1	3
Total costs	3473100	3450900
%GAP	0.70	0.84
%SAVE1	17.97	16.90
%SAVE2	8.92	9.38

(c) Performance measures with respect to c_i^h

	U[5000,15000]	U[1000,19000]
Total costs	3485900	3438100
%GAP	0.71	0.82
%SAVE1	17.95	16.92
%SAVE2	7.50	10.80

(d) Performance measures with respect to $\frac{c_i^{em}}{c_i^{tr}}$

	5	10
Total costs	3377600	3546400
%GAP	0.81	0.72
%SAVE1	17.24	17.63
%SAVE2	9.55	8.76

(e) Performance measures with respect to ET_i^{tr}

	0.1	0.25
Total costs	3361200	3562800
%GAP	0.79	0.74
%SAVE1	18.98	15.89
%SAVE2	8.92	9.38

(f) Performance measures with respect to W^{max}

	0.1	0.25
Total costs	3718400	3205600
%GAP	0.87	0.66
%SAVE1	15.99	18.88
%SAVE2	9.06	9.24

(g) Performance measures with respect to N

	20	50	100
Total costs	1236100	3049100	6100900
%GAP	0.99	0.78	0.53
%SAVE1	17.60	17.31	17.39
%SAVE2	8.91	9.21	9.30

(h) Average results for all instances

Total costs	3462000
%GAP	0.77
%SAVE1	17.43
%SAVE2	9.15

- 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 €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 €50 per hour. Emergency supply lead-time was set at the average level of one day with the costs of €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

Part #	Part title	m_i (day^{-1})	μ_I (day^{-1})	Price (€)	Spare parts inventory levels (S_{i1}, S_{i2}): No pooling	Pooling: ET_i^{tr} =2	Pooling: ET_i^{tr} =4	Pooling: ET_i^{tr} =6
1	flap elec. control	0.0229	0.0141	143450	2, 2	1, 1	1, 1	2, 1
2	flap hydr. control	0.0029	0.0204	113625	0, 0	0, 0	0, 0	0, 0
3	fuel control	0.0143	0.0263	94575	1, 1	1, 1	1, 1	1, 1
4	autopilot computer	0.0343	0.0303	89900	3, 3	2, 2	3, 2	3, 2
5	engine display	0.0114	0.0417	74875	1, 1	1, 1	1, 1	1, 1
6	APU generator gearbox	0.0029	0.0084	55750	0, 0	1, 0	1, 0	0, 0
7	cold air unit	0.0029	0.0119	50292	1, 1	1, 0	1, 0	1, 0
8	multi process unit	0.0400	0.0500	47775	3, 3	3, 2	3, 2	3, 2
9	actuator flap nacelle	0.0029	0.0217	45750	1, 1	1, 0	1, 0	1, 0
10	pitch computer	0.0171	0.0294	45325	2, 2	2, 1	2, 2	2, 2
11	radar	0.0086	0.0345	44100	1, 1	1, 1	1, 1	1, 1
12	roll computer	0.0057	0.0185	43350	1, 1	1, 1	1, 1	1, 1
13	rudder servo	0.0057	0.0167	41325	1, 1	1, 1	1, 1	1, 1
14	TMS computer	0.0400	0.0147	40225	6, 6	5, 4	5, 5	5, 5
15	generator	0.0457	0.0164	37550	6, 6	5, 5	5, 5	5, 5
16	flap actuator outboard	0.0457	0.0357	34800	4, 4	3, 3	4, 3	4, 3
17	roll splr. servo	0.0114	0.0139	33925	2, 2	2, 2	2, 2	2, 2
18	cabin pressure control	0.0086	0.0159	32475	2, 2	2, 1	2, 1	2, 1
19	instr. switching unit	0.0257	0.0500	31425	3, 3	2, 2	2, 2	2, 2
20	prop. control unit	0.0171	0.0125	30750	3, 3	3, 3	3, 3	3, 3
21	distance bearing	0.0143	0.0130	30025	3, 3	3, 2	3, 2	3, 2
22	servo altimeter	0.0400	0.0278	29000	4, 4	4, 3	4, 3	4, 4
23	speed brake hydr. act.	0.0086	0.0417	28700	2, 2	1, 1	1, 1	1, 1
24	navigation selector	0.0171	0.0370	28325	2, 2	2, 2	2, 2	2, 2
25	yaw actuator	0.0114	0.0167	27625	2, 2	2, 2	2, 2	2, 2
26	display processor unit	0.0886	0.0476	27425	6, 6	5, 4	5, 5	5, 5
27	flap indicator switch	0.0029	0.0213	27375	1, 1	1, 0	1, 0	1, 1
28	flap screw jack	0.0171	0.0500	27050	2, 2	2, 1	2, 1	2, 2
29	EFIS display	0.0286	0.0196	26075	4, 4	4, 3	4, 3	4, 3
30	TMS display	0.0200	0.0167	25650	4, 4	3, 3	3, 3	3, 3
31	discharge valve	0.0086	0.0137	25250	2, 2	2, 1	2, 2	2, 2
32	air speed indicator	0.0143	0.0192	25000	3, 3	2, 2	2, 2	2, 2
				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

Alfredsson, P., Verrijdt, J., 1999. Modeling emergency supply flexibility in a two-echelon inventory system. Management Science 45, 1416-1431.

Archibald, T.W., Sassen, S.A., Thomas, L.C., 1997. An optimal policy for a two depot inventory problem with stock transfer. Management Science 43, 173-183.

Axsäter, S., 1990. Modelling emergency lateral transshipments in inventory systems. Management Science 36, 1329-1338.

Axsäter, S., 2003. Evaluation of unidirectional lateral transhipments and substitutions in inventory systems. European Journal of Operational Research 149, 438-447.

Bazaraa, M.S., Sherali, H.D., Shetty, C.M., 1993. Nonlinear Programming: Theory and Algorithms, 2^{nd} ed, Wiley, New York.

Cohen, M.A., Kleindorfer, P.R., Lee, H.L., 1986. Optimal stocking policies for low usage items in multi-echelon inventory systems. Naval Research Logistics 33, 17-38.

Cohen, M.A., Kleindorfer, P.R., Lee, H.L., Pyke, D.F., 1992. Multi-item service constrained (s,S) policies for spare parts logistics systems. Naval Research Logistics 39, 561-577.

Das, C., 1975. Supply and redistribution rules for two-location inventory systems: one-period analysis. Management Science 21, 765-776.

Dowdy, L., Eager, D., Gordon, K., Saxton, L., 1984. Throughput concavity and response time convexity. Information Processing Letters 19, 209-212.

Evers, P.T., 2001. Heuristics for assessing emergency transshipments. European Journal of Operational Research 129, 311-316.

Fisher, M.L., 1985. An applications oriented guide to Lagrangian relaxation. Interfaces 15, 10-21.

Grahovac, J., Chakravarty, A., 2001. Sharing and lateral transshipment of inventory in a supply chain with expensive low-demand items. Management Science 47, 579-594.

Gross, D., 1963. Centralized inventory control in multilocation supply systems. Chapter 3 in Scarf, H.E., Gilford, D.M., and Shelly, M.W. (Eds.), Multistage Inventory Models and Techniques, Stanford University Press, Stanford.

Held, M., Wolfe, P., Crowder, H., 1974. Validation of subgradient optimisation. Mathematical Programming 6, 62-88.

Herer, Y.T., Rashit, A., 1999. Lateral stock transshipments in a two-location inventory system with fixed and joint replenishment costs. Naval Research Logistics 46, 525-547.

Herer, Y.T., Tzur, M., Yücesan, E., 2002. The Multi-location transshipment problem, Working Paper, Faculty of Industrial Engineering and Management, Technion University.

Hoadley, B., Heyman, D.P., 1977. A two-echelon inventory model with purchases, dispositions, shipments, returns and transshipments. Naval Research Logistics 24, 1-19.

Karmarkar, U.S., 1987. The multi-location multi-period inventory problem: bounds and approximations. Management Science 33, 86-94.

Kranenburg, A.A., Van Houtum, G.J., 2003. A multi-item spare parts inventory model with customer differentiation. Working Paper, Technische Universiteit Eindhoven, Forthcoming.

Krishnan, K.S., Rao V.R.K., 1965. Inventory control in N warehouses. Journal of Industrial Engineering 16, 212-215.

Kukreja, A., Schmidt, C.P., Miller, D.M., 2001. Stocking decisions for low-usage items in a multilocation inventory system. Management Science 47, 1371-1383.

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

Needham, P.M., Evers, P.T., 1998. The influence of individual cost factors on the use of emergency transshipments. Transportation Research E 34, 149-160.

Robinson, L.W., 1990. Optimal and approximate policies in multiperiod, multilocation inventory models with transshipments. Operations Research 38, 278-295.

Rudi, N., Kapur, S., Pyke, D., 2001. A two-location inventory model with transshipment and local decision making, Management Science 47, 1668-1680.

Rustenburg, J.W., Van Houtum, G.J., Zijm, W.H.M., 2003. Exact and approximate analysis of multi-echelon, multi-indenture spare parts systems with commonality. Chapter 7 in Shantikumar, J.G., Yao, D.D., Zijm, W.H.M., (Eds.), Stochastic Modeling and Optimization of Manufacturing Systems and Supply Chains, Kluwer, Boston.

Sherbrooke, C.C., 1992a. Optimal Inventory Modeling of Systems. John Wiley & Sons, New York.

Sherbrooke, C.C., 1992b. Multi-echelon inventory systems with lateral supply. Naval Research Logistics 39, 29-40.

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

Tagaras, G., 1999. Pooling in multi-location periodic inventory distribution systems. Omega International Journal of Management Science 27, 39-59.

Tagaras, G., Cohen, M.A., 1992. Pooling in two-location inventory systems with non-negligible replenishment lead times. Management Science 38, 1067-1083.

Thonemann, U.W., Brown, A.O., Hausmann, W.H., 2002. Easy quantification of improved spare parts inventory policies. Management Science 48, 1213-1225.

Timmers, B., 1999. Pooling van repareerbare wisselstukken tussen luchtvaartmaatschappijen. Master Thesis, Centre for Industrial Management, K.U.Leuven.

Van Houtum, G.J., Zijm, W.H.M., 2000. On the relation between cost and service models for general inventory systems. Statistica Neerlandica, 54, 127-147.

Wong, H., Cattrysse, D., Van Oudheusden, D., 2002. Pooling of repairable spare parts in a multi-hub system. Working Paper, Centre for Industrial Management, K.U.Leuven.

Xu, K., Evers, P.T., Fu, M.C., 2003. Estimating customer service in a two-location continuous review inventory model with emergency transhipments. European Journal of Operational Research 145, 569-584.

Yanagi, S., Sasaki, M., 1992. An approximation method for the problem of a repairable-item inventory system with lateral supply. IMA Journal of Mathematics Applied in Business and Industry 3, 305-314.