Journal Pre-proof


# Operational level planning of a multi-item two-echelon spare parts inventory system with reactive and proactive interventions

E. Topan, M.C. van der Heijden

PII: S0377-2217(19)31046-X
DOI: https://doi.org/10.1016/j.ejor.2019.12.022
Reference: EOR 16230

To appear in: *European Journal of Operational Research*

Received date: 14 May 2018
Accepted date: 12 December 2019

Please cite this article as: E. Topan, M.C. van der Heijden, Operational level planning of a multi-item two-echelon spare parts inventory system with reactive and proactive interventions, *European Journal of Operational Research* (2019), doi: https://doi.org/10.1016/j.ejor.2019.12.022

**Highlights**

- We consider operational spare parts planning using real-time information
- We integrate lateral transshipments, emergency shipments, stock allocation
- We consider both reactive and proactive operational interventions
- Our approach is computationally efficient to solve real life problems
- Our experiment with a manufacturers case data shows significant downtime reduction

# Operational level planning of a multi-item two-echelon spare parts inventory system with reactive and proactive interventions

E. Topan[a,] , M.C. van der Heijden[a]

[a]*Industrial Engineering and Business Information Systems (IEBIS), Faculty of Behavioural Management and Social Sciences, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands*

**Abstract**

In this paper, we investigate operational spare parts planning in a multi-item two-echelon distribution system, taking into account real-time supply information in the system. We consider a broad range of operational interventions, either *reactive* (to solve a shortage) or *proactive* (to avoid a shortage). These interventions particularly include lateral transshipments between warehouses (local warehouses), emergency shipments from the depot (central warehouse), and doing nothing and waiting for pipeline inventory. We propose an integrated approach to determine the optimal timing and size of each intervention type to minimize the total downtime and shipment costs associated with interventions. Data from a leading original equipment manufacturer of high-tech systems is used to test the performance of our approach. We find that our integrated approach reduces total downtime considerably with a very limited increase in total shipment costs. Proactive emergency shipments contribute most to downtime reduction. The benefit of our approach is higher for high demand parts. Allowing complete pooling between warehouses increases downtime savings and usage of proactive emergency shipments even further. Our approach is efficient enough to solve practical size problems. We also propose a heuristic based on a greedy algorithm, which is well known in the literature. We find that the gap between the

*Corresponding author

*Email addresses:* e.topan@utwente.nl (E. Topan), m.c.vanderheijden@utwente.nl (M.C. van der Heijden)

heuristic and the optimal solution is relatively large.

*Keywords:* inventory, operational planning, two-echelon, lateral transshipment, emergency shipment

## 1. Introduction

Capital goods are advanced technical systems that are critical for producing services and goods. The downtime penalty costs of these systems are extremely high, e.g., thousands to hundreds of thousands of Euros per hour. Therefore, the availability of spare parts is crucial to reduce downtimes. Spare parts are often provided by the original equipment manufacturers (OEMs).

To reduce downtime, manufacturers employ advanced supply chain networks all around the world and use advanced tactical level spare parts inventory policies. Nevertheless, demand may not be satisfied directly from stock and stockouts may occur. Spare parts planners can still reduce downtime by avoiding stockouts or resolving them as quickly as possible. To do so, planners use various short-term operational interventions, each having different shipment costs and response times, and also real-time on-hand and pipeline stock information which is not available when tactical decisions are being made. Operational planning exploiting such real-time information raises several questions: From which supply location (depot or warehouse) can a part request be best fulfilled when it is out of stock at the warehouse that receives the request. When and how can stocks be moved proactively to reduce stockout risks in the supply chain? Is it sometimes justified to backorder a request and wait for the delivery of pipeline stock that will arrive later? How are these decisions influenced by downtime costs? Foremost, how are these decisions related and how should they be integrated? These are typical questions that planners face every day.

Efficient and effective spare parts planning has received considerable attention in the literature, where most studies focus on tactical planning (Basten & van Houtum, 2014, Hu, Boylan, Chen, & Labib, 2018.) Despite its practical importance, operational planning of spare parts supply has received less attention (Topan, Eruguz, Ma, van der Heijden, & Dekker, 2019). Operational planning differs from tactical planning in three ways: (1) It focuses on the short-term: The decisions are made daily or at short notice to influence the short-term performance. Tactical decision parameters are often

fixed and cannot be changed, e.g, total on-hand and pipeline stock in the supply network. (2) Various sorts of real-time information on the actual state of the supply chain are available for decision making, e.g., number of on-hand and pipeline stock, and delivery times of pipeline stock at each location. (3) The decisions depend on system state at the time of the decision, which itself changes every time these decisions are made. Therefore, the system behaviour is typically characterized by transient behaviour rather than steady state behaviour.

Motivated by practical applications, we propose a model and a solution approach to determine operational level interventions (or actions) in a multi-item two-echelon spare parts supply network. Our primary concern is downtime reduction. Yet, in order to achieve a better balance between responsiveness and cost efficiency, we minimize the total downtime and shipments costs for interventions. We particularly consider stock allocation, lateral transshipments, and emergency shipments from the depot (in other words expedited replenishment) as interventions. As an alterative option, we consider doing nothing and backordering a request and waiting for the delivery of pipeline stock, which we refer to as *waiting for pipeline stock*. We classify interventions in two groups depending on whether the intervention is made before or upon stockout: *proactive* interventions to reduce future stockout risks and *reactive* interventions to fulfill a demand that is not satisfied directly from stock. Moreover, we allow two types of reviews for intervention decisions: planned *periodic reviews* and unplanned *opportunistic reviews* (i.e., stockout events are used as additional review opportunities). Throughout the paper, we use the terms reactive and proactive to identify interventions, and periodic and opportunistic for reviews. We formulate a mixed integer programming (MIP) model to integrate all intervention decisions. We use our MIP model for both review types. Using our MIP model, we integrate stock allocation, rebalancing, demand fulfillment, and emergency supply decisions in a spare parts network in a single model. We consider exact solution of our model and also propose a greedy heuristic to determine interventions.

We use case data provided by a world leading OEM in the semiconductor industry to test the performance of our model. Our main findings are as follows: An experiment with 360 parts with the highest turnover rate reveals that the total downtime can be reduced by one-third without increasing shipment costs significantly. Optimal reactive interventions can be determined fairly well using a simple heuristic, satisfying demand from the nearest supply location with positive stock. Therefore, the downtime reduction achieved

in the experiment is explained mainly by proactive interventions. An experiment based on categorizing items according to demand rate and unit price reveals that the downtime reduction is high for fast movers, and it is almost negligible for slow movers. The downtime reduction is explained predominately by proactive emergency shipments for fast moving expensive parts whereas it is explained almost evenly by both proactive emergency and proactive lateral transshipments for fast moving cheap parts. The average computation time for determining the optimal interventions per item per day is slightly less than half a second. Using the findings of the case study, we develop insights into which intervention types contribute most to downtime reduction, and what kind of spare parts benefit most from downtime reduction when our integrated operational planning approach is used. All these findings are based on exact solution of our problem. Our numerical experiment to test the performance of the greedy heuristic reveals that the greedy heuristic, which performs extremely well in tactical spare parts planning, yields relatively large gap with respect to the exact solution. We attribute this to the myopic behaviour of the greedy heuristic.

Our paper is organized as follows: Section 2 summarizes the related literature. Section 3 describes our problem. In Section 4, we introduce our model and the greedy heuristic. In Section 5, we discuss our numerical experiments and findings. Finally, in Section 6, we draw our conclusions.

## 2. Contribution to the literature

Our paper contributes to four main areas of research: (i) expediting and dual sourcing, (ii) lateral transshipments, (iii) stock allocation, and (iv) spare parts supply planning, particularly operational level planning. Above all, our largest contribution is to operational spare parts planning.

There are several papers on (i) dual sourcing and expediting (e.g., Veeraraghavan & Scheller-Wolf, 2008, Song & Zipkin, 2009 Arts, Basten, & van Houtum, 2016), (ii) lateral transshipments (e.g., Kranenburg & van Houtum, 2009, Paterson, Kiesmüller, Teunter, & Glazebrook, 2011, Paterson, Kiesmüller, Teunter, & Glazebrook, 2012, Glazebrook, Paterson, Rauscher, & Archibald, 2015), and (iii) stock allocation (e.g., van der Heijden, Diks, & de Kok, 1997, Marklund & Rosling, 2012). The major difference from the vast majority of these three streams is three-fold: First, in contrast to most of these papers, which typically focus on tactical planning and investigating impact of expediting, lateral transshipments, and stock allocation on optimal

stock levels, we make operational planning decisions using real-time information of the supply chain. Second, unlike most papers, which assume fixed (expediting, lateral transshipment, or allocation) rules and policies, we do not assume a fixed decision rule and we formulate an MIP to solve the optimal decisions. Third, we integrate emergency shipment (expediting), lateral transshipments, and stock allocation decisions.

There have been several research studies in spare parts supply planning (see Sherbrooke, 2004 Muckstadt, 2005, and van Houtum & Kranenburg, 2015 for books, and Basten & van Houtum, 2014, Hu et al., 2018 for recent reviews). These papers mostly focus on tactical planning, e.g., determining the (near-) optimal inventory policy and finding (near-) optimal inventory parameters. Yet, there are papers on operational level planning of spare parts (see Topan et al., 2019, for a recent review). In Table 1, we compare our paper and the most closely related papers (Caggiano, Muckstadt, & Rappold, 2006, Grahovac & Chakravarty, 2001, Hoadley & Heyman, 1977, Howard, Marklund, Tan, & Reijnen, 2015, and Tiemessen, Fleischmann, van Houtum, van Nunen, and Pratsini, 2013). Our paper differs from these papers particularly in four ways: First, we consider a broad range of proactive and reactive intervention options seen in practice, namely, (i) regular replenishments, (ii) proactive and (iii) reactive emergency shipments from the depot, (iv) proactive and (v) reactive lateral transshipments between locations at the downstream level, (vi) stock allocation, and (vii) backorder clearing (stock allocation during stockouts). Second, we consider waiting for pipeline stock as an option competing against other options rather than as a default option. Third, we clearly distinguish tactical and operational decisions by assuming fixed base stock levels and introducing them to the problem as constraints. Fourth, we consider a general setting with a multi-item, two-echelon and general demand setting with positive lead times. Above all, Caggiano et al. (2006) is the closest paper to ours. Similar to our paper, they consider a multi-item two-echelon setting and propose an MIP formulation (here we refer to their base model as ESAM) to integrate proactive expediting and stock allocation decisions. In addition to these, our paper includes (i) lateral transshipments, and (ii) reactive interventions (to deal with stockouts), and (iii) we consider both periodic reviews and unplanned opportunistic reviews.

Table 1: Papers on operational spare parts planning that are closely related to our paper.

| Related Papers | Intervention | | | | | Intervention type | | Review type | | Problem/model setting | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Expediting (emergency shipments) | Lateral transshipments | Stock allocation | Backorder clearing | Waiting for pipeline stock | Proactive | Reactive | Unplanned, Continuous / opportunistic review | Planned, Periodic review | Number of items | Number of echelons | Planning horizon | Demand | Supply | Tactical parameters (base stock levels) |
| Caggiano et al. (2006) | x | | x | | | x | | | x | Multi | Multi | Finite (rolling) | General distribution | Deter-ministic | Not mentioned |
| Grahovac & Chakravarty (2001) | x | x | | | x | x | x | x | | Single | Multi | Infinite | Poisson | Deter-ministic | Base stock levels also optimized |
| Hoadley & Heyman (1977) | x | x | | | | x | x | | x | Single | Multi | Finite | General distribution | Zero lead time | Not mentioned |
| Howard et al. (2015) | x | x | | x | x | | x | x | | Single | Multi | Infinite | Poisson | Deter-ministic | Base stock levels also optimized |
| Our paper | x | x | x | x | x | x | x | x | x | Multi | Multi | Finite (rolling) | General distribution | Deter-ministic | Base stock levels fixed |

## 3. Problem

We consider a multi-item two-echelon spare parts inventory system consisting of a single depot (central warehouse) and multiple warehouses (local warehouses). Demand for spare parts is random and stationary over time. It arrives only at warehouses, and one at a time. It is independent over parts, locations and mutually exclusive time intervals.

The demand for each part at each warehouse is satisfied directly from stock if there is sufficient stock. Otherwise, this is considered as a *stockout*, and the amount that cannot be satisfied is met by one of the following reactive intervention, or demand fulfillment, options: (i) emergency shipment from the depot, (ii) lateral transhipment from another warehouse, (iii) waiting for pipeline stock until an order in the pipeline arrives at one of the locations (depot or warehouse), which is typically followed by a lateral or emergency shipment.

Apart from reactive interventions, we also allow proactive interventions to reduce downtime. The following proactive intervention options are considered: (i) regular replenishment from the depot to warehouses, (ii) proactive

emergency shipment from the depot to warehouses (with shorter lead time and higher cost), (iii) proactive lateral transhipment between two warehouses.

Proactive interventions, including regular replenishments, are reviewed periodically with a *review period* which is identical for all parts and locations. The review period is taken as unit time (review period is typically one day as we focus on daily planning problems), and each periodic review marks the beginning of the corresponding review period, which we refer to as *periodic review point*. Apart from these fixed review points, we also consider decision moments for reactive interventions as an opportunity to review proactive interventions. We allow for proactive interventions during opportunistic reviews, as some reactive interventions make sense only when combined with a proactive interventions. For example, we may solve a stockout at warehouse A by a reactive lateral transshipment from a nearby warehouse B having only one part, if at the same time we replenish warehouse B by a proactive lateral transshipment from a remote warehouse C.

The replenishments at the depot are controlled according to an echelon stock policy to ensure that the total on-hand and pipeline inventory at all locations are fixed for each part. All proactive interventions, i.e., replenishment of warehouses (regular or emergency) and proactive lateral transshipments, are made centrally by solving an MILP model (the details for this model are explained in Section 4). Hence, we do not assume any fixed policy for warehouses. Yet, (i) each warehouse has an upper bound for its base stock level (typically determined at the tactical planning level); however, warehouses are not necessarily replenished to this level. (ii) Furthermore, proactive emergency shipments are considered to expedite replenishment.

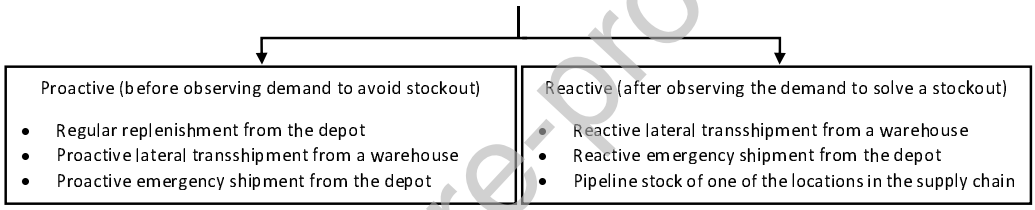
The shipment lead times for proactive and reactive interventions are constant. The shipment lead times for proactive interventions are integer multiples of the review period (this is not necessary for reactive interventions). This is not constraining in practice if we choose the review period short enough (e.g., a day). Since lead times are integer multiples of review periods and proactive interventions that are determined at an opportunistic review are executed at the next periodic review point, all proactive shipments are ordered and received at periodic review points.

Interventions includes the following costs: (i) a unit shipment cost, if a shipment is involved, and (ii) a unit time backorder cost (downtime penalty charge), if demand cannot be satisfied directly from stock. This cost is charged for downtimes associated with waiting for pipeline stock as well as the downtimes during reactive interventions, i.e., reactive shipment lead time,

and if the delivery is from pipeline stock, plus the remaining time until the delivery of pipeline stock to that location. We assume that pipeline process cannot be expedited.

Apart from being made upon stockouts instead of before, reactive interventions differ from proactive interventions also because they are more expensive (as they involve additional downtime during shipment) and more urgent (when a part is available, a reactive intervention is made immediately; whereas a proactive intervention can wait, e.g., until the next periodic review). All proactive and reactive intervention options considered in our system are summarized in Figure 1.

Figure 1: Spare parts interventions in our two-echelon supply network.




The sequence of events in each period is as follows:

1. At the periodic review point, pipeline orders that are due, arrive at their destination locations. Using these shipments, backorders are cleared (if there are any). Decisions for new proactive interventions at all warehouses and replenishment orders at the depot are made, and their shipment orders are placed.

2. Demand arrives at warehouses throughout the period. It is satisfied immediately either from stock or by one of the reactive demand fulfillment options. Downtime is incurred during the shipment lead time for the reactive intervention. Reactive decision points are used as an opportunity to review proactive interventions.

3. The on-hand and pipeline stocks, and backordered units at the end of the period are updated. Downtime costs are incurred for backorders at the end of the period.

Our primary concern is to make joint proactive and reactive intervention decisions to minimize total downtime. Yet, we do not want to minimize

downtime at any cost. Therefore, our objective is to minimize the total shipment and downtime costs. We do not include holding cost in the objective function since the total maximum inventory at the tactical level has been determined at the tactical level and is therefore fixed for each part.

## 4. Decision model

### 4.1. Model outline

We propose a mixed integer programming (MIP) formulation for our proactive and reactive intervention decisions. We test our planning logic in a discrete event simulation. In the simulation, we consider two types of reviews:

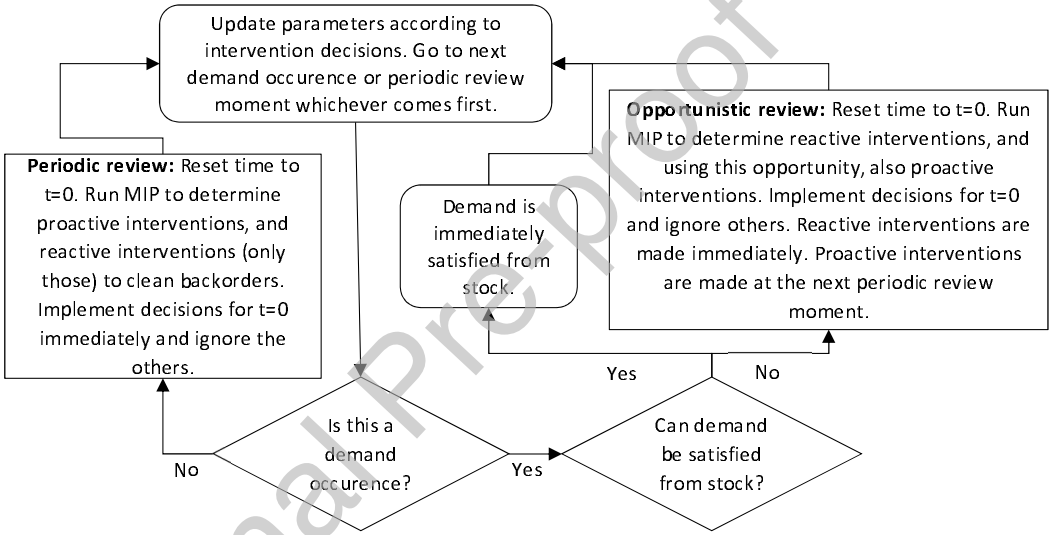
- *periodic review*, at the beginning of each period, to make proactive intervention decisions, and if there are any, using pipeline orders that arrive at this periodic review point, to make reactive intervention decisions to clear backorders,
- *opportunistic review*, at stockout events, to make reactive shipment decisions for stockouts, and using this opportunity, to make proactive shipment decisions to balance the inventory in the network. During each period, this is run as many times as the number of stockouts.

In both review types, (i) we call exactly the same MIP model to make our intervention decisions, (ii) we run the model on a rolling horizon, and (iii) we implement only the decisions for the initial period. Yet, we distinguish between periodic reviews and opportunistic reviews for three reasons: First, review times of periodic reviews are fixed and known, whereas opportunistic reviews are random and exact times are unknown. Second, the reactive intervention decisions made (by the MIP model) at opportunistic reviews are to fulfill demand from stocks (of other warehouses and the depot), whereas the reactive intervention decisions made (by the MIP model) at periodic reviews use pipeline stocks (of all warehouses and the depot) to clear backorders. Third, we observe this distinction also in practice, e.g., daily or weekly reviews vs. reviews triggered by exception messages.

All reactive interventions determined by the MIP model are placed immediately. The proactive interventions that are determined periodically at periodic review points are ordered immediately. The proactive interventions that are determined at opportunistic review points are ordered at the next

periodic review point since proactive intervention orders are only allowed at periodic review points. In all these cases, the shipments associated with interventions are received after the corresponding lead time. When a demand is satisfied directly from stock, this does not require running the MIP. We assume that demand does not arrive exactly at periodic review points. An overview of our integrated model is illustrated in Figure 2. Our MIP model

Figure 2: Flow diagram for our integrated model.




and the greedy heuristic are explained in Sections 4.2 and 4.3, respectively.

### 4.2. Model and its exact solution

The MIP model is based on a discrete time model with a fixed planning horizon length $T$. Without loss of generality, we define the start of the first period in an MIP run as $t = 0$. When the MIP model is triggered by an opportunistic review, we take the next review period as time zero in our MIP model. Following a rolling horizon procedure, the decisions for the initial period, at $t = 0$ for interval [0,1), are put into effect immediately, and those for subsequent periods are ignored since this procedure is repeated for each period. The MIP model assumes that the future demand that cannot be satisfied from stock is backordered. Note that when a stockout occurs,

our integrated approach actually invokes our MIP model. According to the solution of the model, the demand is responded by a reactive intervention, and it is either satisfied from other locations or backordered (and satisfied from pipeline stock). Full backordering assumption, which we use only for the MIP formulation, facilitates using the information about cumulative inventory and demand at warehouses until a specific time as input. Therefore, for each part $i$ and each location $l$ and $t \quad \{0, 1, \ldots, T-1\}$, we let $\bar{S}_{ilt}$ denote the known cumulative supply of parts until the beginning of period $t$. This includes initial inventory on-hand plus all proactive intervention orders in the pipeline that arrives before $t$. We define $D_{ijt} = D_{ij}(t, t+1)$ to denote the demand for part $i$ that arrives at warehouse $j$ in interval $[t, t+1)$ and $D^{cum}_{ijt} = \sum_{l=0}^{t} D_{ijl}$ to denote the cumulative demand for part $i$ at warehouse $j$ during $(0, t+1)$.

For each part $i$, warehouse $j$ and time $t$, the model determines the following proactive intervention decisions: (i) the number of regular shipments, $y^{preg}_{ijt}$, allocated by the depot, (ii) the number of proactive lateral transshipments from warehouse $m$, $y^{plat}_{imjt}$, (iii) the number of proactive emergency shipments from the depot, $y^{pem}_{ijt}$. Furthermore, for each part $i$ and warehouse $j$, we let $B_{ij}$ denote the number of known backorders at time $t = 0$. This amount can be satisfied by reactive lateral transshipments and emergency shipments. Therefore, we let $y^{rlat}_{imj}$ denote the number of reactive lateral transshipments from warehouse $m$ and $y^{rem}_{ij}$ denote the number of reactive emergency shipments from the depot. By definition, reactive interventions cannot be planned for future periods. Therefore, reactive interventions are defined only for the initial period. Together with the known cumulative supply $\bar{S}_{ijt}$, including shipments associated with all (proactive and reactive) interventions add up to the cumulative inventory $S_{ijt}$ for period $t$. The fixed cumulative inventory levels are updated according to the decisions made for $t = 0$ after each run of the model. For each part $i$ and warehouse $j$, $B_{ij}$ is updated according to changes in real-time backorders: It increases by 1 when a demand arrives at the warehouse, and decreases when this demand is satisfied (from direct stock or by one of the reactive interventions), and remains the same when the demand is backordered. The known backorders and known cumulative supply are fixed inputs for the MIP. Our notation for all parameters and variables is summarized in Table 2.

The mixed integer programming model formulation of our problem is

Table 2: Notation for the MIP model

| | |
|---|---|
| **Input parameters** | |
| $T$ | Time horizon |
| $t$ | Discrete time index, $t = 0, \ldots, T - 1$ |
| $i$ | Part index, $i = 1, \ldots, I$ |
| $j$ | Warehouse index, $j = 1, \ldots, J$ |
| $l$ | Location index extended to include the depot, $l = 0, \ldots, J$ |
| $C_j$ | Set of all warehouses except for warehouse $j$, $\{k \vert 1 \leq k \leq J, k = j\}$ |
| $\lambda_{ij}$ | Demand rate for part $i$ at warehouse $j$ |
| $L_i$ | Regular replenishment lead time of part $i$ to the depot |
| $L_{ij}^{preg}$ | Regular replenishment lead time of part $i$ to warehouse $j$ |
| $c_{ij}^{preg}$ | Unit regular replenishment cost of part $i$ to warehouse $j$ |
| $L_{ij}^{pem}$ | Proactive emergency shipment lead time of part $i$ from the depot to warehouse $j$ |
| $c_{ij}^{pem}$ | Unit proactive emergency shipment cost of part $i$ from the depot to warehouse $j$ |
| $c_{imj}^{plat}$ | Proactive lateral shipment cost of part $i$ to warehouse $j$ from warehouse $m \quad C_j$ |
| $L_{imj}^{plat}$ | Proactive lateral shipment time of part $i$ to warehouse $j$ from warehouse $m \quad C_j$ |
| $L_{imj}^{rlat}$ | Reactive lateral transshipment lead time from warehouse $m \quad C_j$ to warehouse $j$ for part $i$ |
| $c_{imj}^{rlat}$ | Unit reactive lateral transshipment cost from warehouse $m \quad C_j$ to warehouse $j$ for part $i$ |
| $L_{ij}^{rem}$ | Reactive emergency lead time from the depot to warehouse $j$ for part $i$ |
| $c_{ij}^{rem}$ | Unit reactive emergency shipment cost from the depot to warehouse $j$ for part $i$ |
| $p_{ij}$ | Downtime penalty charge (backorder cost) per unit time for part $i$ at warehouse $j$ |
| $s_{ij}$ | Base stock level for part $i$ at warehouse $j$ |
| $s_{i0}$ | Echelon base stock level for part $i$ at the depot |
| $\bar{S}_{ilt}$ | Known cumulative supply of part $i$ at location $l$ until the beginning of period $t$ |
| $B_{ij}$ | The number of known backorders for part $i$ at warehouse $j$ |
| **Random parameters** | |
| $D_{ijt}$ | Demand for part $i$ at warehouse $j$ in period $t$ |
| **Decision variables** | |
| $y_{ijt}^{preg}$ | Number of regular replenishments of part $i$ to warehouse $j$ in period $t$ |
| $y_{imjt}^{plat}$ | Number of proactive lateral transshipments of part $i$ from warehouse $m$ to warehouse $j$ in period $t$ |
| $y_{ijt}^{pem}$ | Number of proactive emergency shipments of part $i$ from the depot to warehouse $j$ in period $t$ |
| $y_{imj}^{rlat}$ | Number of reactive lateral transshipments of part $i$ from warehouse $m \quad C_j$ to warehouse $j$ in period zero |
| $y_{ij}^{rem}$ | Number of emergency shipments of part $i$ from the depot to warehouse $j$ in period zero |
| **Dependent variables** | |
| $S_{ijt}$ | Cumulative supply of part $i$ at warehouse $j$ until the beginning of period $t$ (including shipments for proactive interventions placed in period $t$) |
| $D_{ijt}^{cum}$ | Cumulative demand for part $i$ at warehouse $j$ until the end of period $t$ |

stated by

$$
\begin{aligned}
Min \quad & \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{t=0}^{T-1} p_{ij}E[(D_{ijt}^{cum}+B_{ij}-S_{ijt})^{+}] + \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{t=0}^{T-1} c_{ij}^{preg} y_{ijt}^{preg} \\
& + \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{t=0}^{T-1} c_{ij}^{pem} y_{ijt}^{pem} + \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{m\ C_j}\sum_{t=0}^{T-1} c_{imj}^{plat} y_{imjt}^{plat} \\
& + \sum_{i=1}^{I}\sum_{j=1}^{J} c_{ij}^{rem} y_{ij}^{rem} + \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{m\ C_j} c_{imj}^{rlat} y_{imj}^{rlat} \\
& + \sum_{i=1}^{I}\sum_{j=1}^{J} p_{ij} L_{ij}^{rem} y_{ij}^{rem} + \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{m\ C_j} p_{ij} L_{imj}^{rlat} y_{imj}^{rlat} \qquad (1)
\end{aligned}
$$

$$
s.t. \quad \bar{S}_{i0t} \geq \sum_{j=1}^{J}\sum_{l=0}^{t} \left(y_{ijl}^{preg} + y_{ijl}^{pem}\right) + \sum_{j=1}^{J} y_{ij}^{rem} \text{ for } i \text{ and } t \qquad (2)
$$

$$
\begin{aligned}
S_{ijt} = \bar{S}_{ijt} + & \sum_{l=0}^{t-L_{ij}^{preg}} y_{ijl}^{preg} + \sum_{l=0}^{t-L_{ij}^{pem}} y_{ijl}^{pem} - \sum_{l=0}^{t}\sum_{m\ C_j} y_{ijml}^{plat} + \sum_{m\ C_j}\sum_{l=0}^{t-L_{imj}^{plat}} y_{imjl}^{plat} \\
& + y_{ij}^{rem} + \sum_{m\ C_j} y_{imj}^{rlat} - \sum_{m\ C_j} y_{ijm}^{rlat} \text{ for } i, j, \text{ and } t \qquad (3)
\end{aligned}
$$

$$
\bar{S}_{ijL_{ij}^{preg}} + y_{ij0}^{preg} + y_{ij0}^{pem} - \sum_{m\ C_j} y_{ijm0}^{plat} + \sum_{m\ C_j} y_{imj0}^{plat} \leq s_{ij} \text{ for } i, j \qquad (4)
$$

$$
y_{ij}^{rem} + \sum_{m\ C_j} y_{imj}^{rlat} - \sum_{m\ C_j} y_{ijm}^{rlat} \leq B_{ij} \text{ for } i, j \qquad (5)
$$

$$
\begin{aligned}
& y_{ijt}^{preg}, y_{ijt}^{pem}, S_{ijt} \quad \mathbb{N}_0 \text{ for } i, j, \text{ and } t \\
& y_{imjt}^{plat} \quad \mathbb{N}_0 \text{ for } i, j, m \quad C_j, \text{ and } t \\
& y_{ij}^{rem} \quad \mathbb{N}_0 \text{ for } i, \text{ and } j \\
& y_{ijm}^{rlat} \quad \mathbb{N}_0 \text{ for } i, j \text{ and } m \quad C_j
\end{aligned}
$$

In the MIP formulation, the objective is to minimize total backorder, regular shipment, proactive emergency shipment, and proactive lateral transshipment costs, reactive emergency shipment costs, backorder (downtime) cost associated with reactive shipments. Constraint (2) states that the total cumulative (regular, proactive emergency, reactive emergency) shipments from the depot at any time $t$ is limited by the known cumulative supply at the

depot $\bar{S}_{i0t}$ for each item $i$. Constraint (3) defines the cumulative supply (including shipments for proactive and reactive interventions) $S_{ijt}$ for each item $i$, warehouse $j$, and time $t$. In constraints (2) and (3), cumulative inventory includes reactive interventions only at t=0 because reactive interventions are allowed only for t=0 to fulfill real (time) demand or known backorders. Constraint (4) guarantees that the inventory position after including proactive interventions at time $t = 0$ does not exceed the tactical local base-stock level $s_{ij}$ for each part $i$ and warehouse $j$. Constraint (4) is necessary to avoid that tactical decisions on base stock levels are violated. Constraint (5) guarantees that reactive interventions are made only for known backorders.

Apart from proactive intervention decisions, the MIP model seeks reactive interventions to resolve backorders (positive $B_{ij}$). Reactive interventions are immediately coupled with backorders. When a demand finds a match with one of the supply options, one unit is deducted from both sides of constraint (5), i.e., supply and demand mathematically cancels out each other. This also explains why shipment lead times for reactive interventions do not appear in constraint (3). Yet, actual demand fulfillment is accomplished after the fixed shipment lead time. Therefore, the cost associated with the shipment, including downtime, is charged in the objective function. With this, we are able to model reactive and proactive interventions in the same model. Since we run the MIP model for each stockout event in a timely order, and all shipment lead times are constant, shipments for interventions do not overlap, e.g., an intervention decision made later cannot fulfill a demand that occurs earlier in the same period. The MIP model is decomposable by parts. Therefore, we solve it for each part $i$ separately. We do not consider any specialized exact solution approach. We use CPLEX 12.6.3 to solve the MIP model exactly.

### 4.3. Greedy heuristic

The greedy heuristic is known for its good performance for tactical spare parts inventory planning (Kranenburg and van Houtum, 2009, and Topan, Bayındır, and Tan 2017). To construct our greedy heuristic, we start with setting all decision variables to zero. At each iteration, we compute the cost reduction achieved by increasing each decision variable and select the decision variable with the highest cost reduction to increase by 1. The procedure is repeated until no cost reduction is possible. To explain the details, we let $\mathbf{S}$ be the vector of cumulative supply $S_{ijt}$ for all parts $i$, warehouses $j$ and periods $t$. Further let $Z(\mathbf{S})$ be the total cost (1) for the cumulative supply

vector $\mathbf{S}$. Let $\Delta_x Z(\mathbf{S})$ be the cost reduction associated with increasing $x$ by 1, where $x$ represents any decision variable in our MIP. Then, the cost reductions for decision variables are expressed by

$$\Delta_{y^{preg}_{ijt}} Z(\mathbf{S}) = c^{preg}_{ij} - p_{ij} \sum_{l=t+L^{preg}_{ij}}^{T-1} P(D^{cum}_{ijl} \geq S_{ijl} + 1 - B_{ij}), \tag{6}$$

$$\Delta_{y^{pem}_{ijt}} Z(\mathbf{S}) = c^{pem}_{ij} - p_{ij} \sum_{l=t+L^{pem}_{ij}}^{T-1} P(D^{cum}_{ijl} \geq S_{ijl} + 1 - B_{ij}), \tag{7}$$

$$\begin{aligned} \Delta_{y^{plat}_{imjt}} Z(\mathbf{S}) &= c^{plat}_{imj} - p_{ij} \sum_{l=t+L^{plat}_{imj}}^{T-1} P(D^{cum}_{ijl} \geq S_{ijl} + 1 - B_{ij}) \\ &+ p_{im} \sum_{l=t}^{T-1} P(D^{cum}_{iml} \geq S_{iml} - B_{im}) \end{aligned} \tag{8}$$

$$\Delta_{y^{rem}_{ij}} Z(\mathbf{S}) = c^{rem}_{ij} + p_{ij} L^{rem}_{ij} - p_{ij} \sum_{l=t}^{T-1} P(D^{cum}_{ijl} \geq S_{ijl} + 1 - B_{ij}), \tag{9}$$

$$\begin{aligned} \Delta_{y^{rlat}_{imj}} Z(\mathbf{S}) &= c^{rlat}_{imj} + p_{ij} L^{rlat}_{imj} - p_{ij} \sum_{l=t}^{T-1} P(D^{cum}_{ijl} \geq S_{ijl} + 1 - B_{ij}) \\ &+ p_{im} \sum_{l=t}^{T-1} P(D^{cum}_{iml} \geq S_{iml} - B_{im}). \end{aligned} \tag{10}$$

We derive equation (8) as follows: Increasing $y^{plat}_{imjt}$ by 1 increases the shipment cost by $c^{plat}_{imj}$. This increases cumulative supply at $j$ starting from time $t+L^{plat}_{imj}$ by an amount 1. It is easy to show that

$$\begin{aligned} \Delta_{S_{ijt}} E[(D^{cum}_{ijt} + B_{ij} - S_{ijt})^+] &= E[(D^{cum}_{ijt} + B_{ij} - (S_{ijt} + 1))^+] \\ &- E[(D^{cum}_{ijt} + B_{ij} - S_{ijt})^+] \\ &= -P(D^{cum}_{ijt} \geq S_{ijt} + 1 - B_{ij}). \end{aligned} \tag{11}$$

Therefore, increasing $y^{plat}_{imjt}$ by 1 decreases expected downtime at $j$ by an amount $\sum_{l=t+L^{plat}_{imj}}^{T-1} P(D^{cum}_{ijl} \geq S_{ijl} + 1 - B_{ij})$. Similarly, increasing $y^{plat}_{imjt}$ by 1 decreases cumulative supply at $m$ starting from time $t$ by an amount 1.

Then, using (11), we write

$$\begin{aligned} E[(D^{cum}_{imt} + B_{im} - (S_{imt} - 1))^+] \; &- \; E[(D^{cum}_{imt} + B_{im} - S_{ijt})^+] \\ &= \; -P(D^{cum}_{imt} \geq S_{imt} - B_{im}). \end{aligned}$$

Therefore, increasing $y^{plat}_{imjt}$ by 1 increases expected downtime at $m$ by an amount $\sum_{l=t}^{T-1} P(D^{cum}_{iml} \geq S_{iml} - B_{im})$. Equations (6) and (7) follow from (8) with one exception: The shipment to warehouse $j$ is from the depot and the depot does not contribute to the total cost in (1), and therefore, the last term in (8) drops. Equations (9) and (10) also follows from (8). Yet, (9) and (10) include additional downtime cost during lead times for reactive interventions $p_{ij}L^{rem}_{ij}$ and $p_{ij}L^{rlat}_{imj}$, respectively.

Although the greedy heuristic is simple, the computational requirement can still be high. At each iteration, the expressions (6)-(8) need to be evaluated for each part $i$, warehouse $j$, and time period $t$, and (9) and (8) for each part $i$, warehouse $j$. As an alternative, we consider early truncation of the greedy algorithm. To do so, for each part $i$, we truncate the decision space at each iteration by only considering the decisions for periods $t < max_j\{L_{ij}\}$. Truncation is reasonable because the impact of interventions made in later periods has often marginal impact on the overall performance and the decisions made for later periods are not executed at all due to the rolling horizon procedure. The impact of truncation is analyzed in Section 5.

## 5. Numerical study

We conduct numerical studies to investigate the value of using our integrated operational planning model. We discuss the experimental design in Section 5.1 and the corresponding numerical results in Section 5.2. We also examine the impact of parameters and the backorder cost estimation method on the value of using our model (Section 5.3), explore the solution quality of the greedy algorithm and the impact of truncation (Section 5.4), and finally investigate the solution quality and the computational efficiency of our model for practical size problems (Section 5.5). Except Section 5.4, we present results that are based on the exact solution of the MIP model.

### 5.1. Experimental design and case data

We use the data provided by the OEM as experimental data. We select their case because it is quite a good representation of spare parts applications.

Furthermore, our study is a part of an ongoing research project on after-sales service of capital goods, in which the OEM has a key role. The OEM is a world leading manufacturer in the semiconductor industry. To ensure spare parts availability to its customers for more than 6000 parts, it operates a global supply network with more than 30 warehouses and one depot. It replenishes spare part stocks according to a base stock policy. The values of the optimal base stock levels are determined by using a multi-item two-echelon tactical inventory planning model, allowing reactive lateral transshipments between warehouses with partial pooling and reactive emergency shipments from the depot. The model is based on Kranenburg and van Houtum (2009) and its extension by van Aspert (2015). Although the planning model is a continuous review model, the company places regular replenishment orders periodically with a review period of 1 day. The real-time demand fulfilment (reactive planning) is performed according to the principle that the part requests that cannot be met directly from stock are fulfilled from the nearest location having that item on stock. We consider the OEM's inventory policy as a benchmark and we use its base stock levels as a tactical upper bound on inventory levels as in equation (4).

The case data include the values of average demand rates, unit prices, base stock levels, and shipment lead times and costs for its parts and global network. Most parts are slow moving and expensive. For the sake of confidentiality, we do not reveal the data, yet we give a reasonable indication of the data. The maximum unit price of parts is expressed in millions of Euros and the maximum demand rate is expressed in hundreds per year. Demand rates and unit prices are highly asymmetric. When parts are ranked according to the demand rate, 20% of the parts represent approximately 86% of the total demand rate of parts. When parts are ranked according to unit price, 20% of the parts represent approximately 93% of the unit price of parts. Figure 3 summarizes the distribution of demand rates and the unit prices of parts in the experiment on a unitless scale (the points indicated in red will be explained later in Section 5.2). The maximum, minimum, and average values of the lead times and the shipment costs of regular replenishment, proactive and reactive emergency shipments and proactive and reactive lateral shipments of the parts considered in the experiment are summarized in Table 3.

By consulting the inventory planners at the manufacturer, we consider the following setting for the remaining parameters: The regular replenishment costs $c_{ij}^{preg}$ are set to €200 for each combination of item $i$ and warehouse

Figure 3: Demand rates and unit prices in the case data.

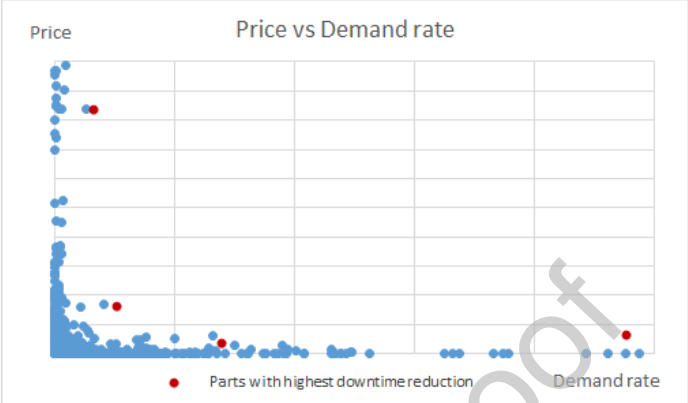



Table 3: Parameters values in the base case setting of the first experiment

| **Parameters** | **Min.** | **Avg.** | **Max.** |
|---|---|---|---|
| $L_{ij}^{preg}$ (days) | 8 | 18 | 33 |
| $c_{ij}^{pem}$, $c_{i0j}^{rem}$ (€/unit) | 300 | 1123 | 1200 |
| $L_{ij}^{pem}$, $L_{i0j}^{rem}$ (hours) | 24 | 48 | 72 |
| $L_{i0}^{preg}$ (days) | 22 | 102 | 386 |
| $c_{ijm}^{plat}$, $c_{ijm}^{rplat}$ (€/unit) | 40 | 316 | 1143 |
| $L_{ijm}^{plat}$, $L_{ijm}^{rplat}$ (hours) | 3 | 17 | 36 |

$j$. Since the case study is based on the global network of the manufacturer, the warehouses are located in different countries and even on different continents. The lateral transshipments are allowed only among warehouses in the same country due to limitations imposed by customs regulations. A downtime penalty charge is either explicitly stated in service contracts or implied by service targets, e.g., aggregate target waiting time, aggregate fill rate etc. Considering a conservative estimate for the OEM's case, we set the downtime penalty charge $p_{ij}$ to €2500 per hour for each part $i$ and warehouse $j$. The planning horizon length $T$ of the MILP model is set to the maximum regular replenishment lead time of each part. We take the review period (for periodic reviews) as 1 day. This duration is long enough to keep order placement manageable in practice (the review period at the OEM is also 1

day), and short enough to make timely and cost effective decisions (see Section 5.3). The OEM does not currently use lateral and emergency shipments options proactively, and it does not distinguish between reactive and proactive interventions. Therefore, lead times and shipment costs are the same for proactive and reactive interventions. Yet, in Section 5.3, we investigate the impact of the review period, planning horizon, downtime penalty charge, and shipment costs and shipment lead times of proactive interventions, and using a complete pooling strategy.

The current policy used by the OEM involves only regular replenishments. In that respect, the current policy can be considered as a special case of our integrated approach, the proactive planning without proactive lateral transshipments and proactive emergency shipments. We consider four variants of our integrated approach:

- *Current policy* excludes proactive lateral transshipments, and proactive emergency shipments. Replenishments of warehouses are filled first-come, first-served by the depot. Reactive intervention decisions are made based on a fixed rule: (1) Demand that cannot be satisfied from stock is satisfied from the nearest location with positive stock. (2) If there is no stock at any location, the request is satisfied from pipeline stock that is first available (hence backorder clear ing decision is fixed at the beginning) and demand is backordered until the pipeline stock arrives. This represents the current policy of the OEM. We use the current policy as our main benchmark.

- *Integrated policy* includes proactive lateral transshipments, proactive emergency shipments, regular replenishment, and reactive interventions. It corresponds to our approach described in Section 4.

- *Integrated without proactive lateral shipments* excludes proactive lateral transshipments, includes proactive emergency shipments, regular replenishment, and reactive interventions. We use this policy as a benchmark to test the impact of lateral transhipments in the integrated policy.

- *Current policy with proactive emergency shipments* extends the current policy by allowing proactive emergency shipments. We use a dual-index policy (Veeraraghavan & Scheller-Wolf, 2008), which uses a second base

stock level for proactive emergency shipments. When inventory position during proactive emergency shipment time drops below this second base stock level, a proactive emergency order is placed to raise inventory position during proactive emergency shipment time to the second base stock level. We use this policy to test the impact of using a proactive emergency shipments using a simple heuristic.

- *Integrated policy without reactive interventions (hence also without opportunistic reviews)* includes proactive lateral transshipments, proactive emergency shipments, regular replenishment by optimal stock allocation. Yet, it excludes reactive interventions. This means, when demand cannot be satisfied directly from stock, it is backordered until the next periodic review point and then satisfied by backorder clearing. We use this policy as a benchmark to test the impact of reactive interventions in the integrated policy.

We note that the comparison with the current policy gives also an insight into a possible comparison with Kranenburg and van Houtum (2009) since base stock policies and the reactive planning assumptions in the current policy are based on this paper. Similarly, the comparison against the integrated without lateral transshipments and the integrated without reactive interventions give some insights about a possible comparison with Caggiano et al. (2006) since our paper differs from their paper in these aspects. For all parts and locations, we determine the dual-index parameters by setting it to the minimum inventory level that satisfies a newsboy ratio of 0.95 by considering the demand distribution during proactive emergency shipment lead time.

Our key performance measure is the total downtime. Yet, we acknowledge that our findings are similar when we base our findings on the total cost. This is attributed the fact that backorder (downtime) costs are very large while shipment costs are small compared to downtime costs and that the total cost is mostly determined by the downtime cost. To evaluate the performance of policies, we use discrete event simulation. Our simulation runs according to the flow in Figure 2 in Section 4.1. All locations start each replication with stock equal to base stock levels. The warm up period and the simulation length (excluding warmup period) are both half a year. 5 replications are simulated for each instance. We consider the average CPU time per day (time it takes to make all decisions for all selected parts during a day) to test the computational performances.

To investigate for which part group operational interventions are most interesting, we identify 4 groups of 15 parts: (i) high-demand and high-price, (ii) high-demand and low-price, (iii) low-demand and high-price, and (iv) low-demand and low-price items. The high-demand and low-demand items are identified by demand rates $> 6$ and $< 0.5$ (units per year), respectively; and high-price and low-price items are identified by unit prices $>$e25000 and $<$e1000, respectively. To select parts for each of the 4 groups, we rank parts according to turnover rate (just like in standard ABC classification, price times demand rate) and then select the top 15 parts each. According to this selection criterion, the selected 15 parts for the high-demand high-price group represents 50% of the total turnover for all parts. Half a year simulation length could be small for low-demand parts. However, a possible experimental error for total cost for these low-demand parts would also be very small for the same reason. To keep the balance between total computation time and accuracy of our results, we take the simulation length as 1 year.

### 5.2. Value of our integrated approach and properties of the optimal policy

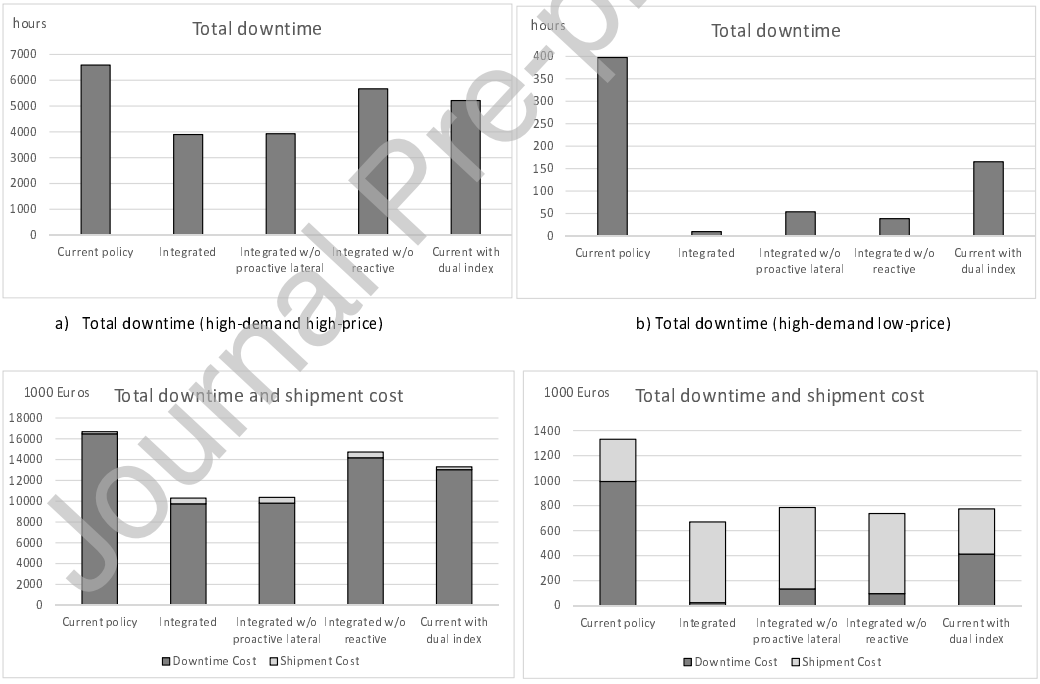
Figure 4 illustrates the total downtime and the distribution of total cost over downtime and shipment costs, respectively. The figures summarize the results for 5 different policies for high-demand high-price (in a and c) and high-demand low-price parts (in b and d). Each value in Figure 4 (also in all figures in the rest of the paper) represents the average of 5 replications for 15 parts over a simulation length of half a year. Based on these two figures, we make the following observations:

- *The downtime reduction by using our integrated approach is high (only) for high-demand parts:* Our integrated approach reduces the (average) total downtime (over half a year) from 6589 to 3898 hours for high-demand and high-price parts (Figure 4a), and from 387 to 10 hours for high-demand and low-price parts (Figure 4b). Downtime savings are higher for high-demand parts, in absolute terms for high-price items and in percentage values for low-price parts, making these two groups of parts interesting for operational planning. Intuitively, proactive emergency shipments and proactive lateral shipments contribute more when there is more inventory at warehouses. As an illustration, we indicate the parts that benefit most from our approach in red in Figure 3. These parts belong to the high-demand high-price group. The downtime reduction for the parts constitute

81% of the total downtime reduction for 15 high-demand high-price parts. For low-demand high-price and low-demand low-price parts, the downtime savings are negligibly small. Therefore, we drop these two groups from further discussions.

- *The downtime reduction is achieved without any significant cost increase:* Figures 4c illustrates that the increase in total shipment cost is relatively small (a few hundred thousand Euros) compared to the total downtime savings (expressed in millions of Euros) for high-demand high-price group. This is attributed to high downtime costs. Our observation is valid, but to a lesser extent, for high-demand low-price group (see Figure 4d).

Figure 4: Total downtime and distribution of total cost for different part groups under different policy settings.




- *The downtime reduction is mainly attributed to the proactive interventions:* We acknowledge that determining reactive interventions based on

distances performs quite well. Using the reactive interventions that are suggested by the MIP solution has almost no impact on total downtime reduction. This is attributed to that downtime cost are very high compared to shipment costs. Therefore, the downtime reduction by using our integrated method (in comparison to the current policy) is attributed mainly to optimal proactive interventions.

- *Leaving out reactive interventions and solving stockouts at periodic review points leads to high downtime*. As seen in Figure 4a, the integrated approach without reactive interventions may perform poorly with respect to integrated policy (yet it is better than the current policy). This is due to the fact that it is very costly to wait until the first periodic review point for a stockout event, e.g., even for a review period of 1 day, waiting for a reactive intervention until the next review, which is on the average half day (12 hours), costs 30000 Euros extra.

- *For high-demand high-price parts, most of the downtime reduction by the optimal proactive intervention decisions is attributed to the proactive emergency shipments:* Including proactive lateral transshipments does not influence the total downtime (and also the total cost) of the optimal solution for high-demand high-price parts, e.g., the total downtime is 3927 when proactive lateral transshipments are excluded, in comparison to 3898 when they are included. In line with this observation, for high-demand high-price parts, a large portion (72%) of the proactive interventions is emergency shipments whereas only a small portion (10%) is lateral transshipments (and the rest is regular replenishments). Our finding is attributed to the lead time reduction enabled by proactive emergency shipments, e.g., according to Table 3, proactive emergency shipments can reduce replenishment lead time from 18 days to 2 days on the average. In contrast, proactive lateral transshipments contribute less to lead time (hence expected downtime) reduction. High holding cost encourages holding more inventory at the depot and less at warehouses (note that the total inventory is fixed in our setting). In our experiments, for high-demand high-price parts, 80% of the total inventory is kept at the depot. This makes lead time (hence expected downtime) reduction by using proactive emergency shipments interesting for these parts.

- *For high-demand low-price parts, the downtime reduction by the optimal proactive interventions is explained by both proactive emergency shipments*

*and proactive lateral transshipments:* In contrast to our observation for high-demand high-price parts, the difference in downtime between options with and without lateral transshipment is larger for high-demand low-price parts. The total downtime is 56 when proactive lateral transshipments are excluded in comparison to 13 when they are included. The proportions of proactive emergency shipments and proactive lateral transshipments are 25% and 10%, respectively. Our findings indicate that the added value of proactive lateral transshipments is much higher for high-demand low-price parts. In contrast to high-demand high-price parts, for high-demand low-price parts, a larger amount, i.e., 43%, of the total inventory is kept at warehouses. This makes proactive lateral transshipments more (proactive emergency shipments less) interesting for high-demand low-price parts.

- *Current policy with dual-index policy for proactive emergency shipments does not perform as well as our integrated approach.* Figure 4a shows that this policy yields poor results relative to our approach. This indicates that unlike reactive intervention decisions, finding a policy for proactive interventions that is based on simple rules and has a good performance may not be easy.

- *Despite the high downtime penalty charge, stockouts may still be resolved by backordering and waiting for pipeline stock:* For high-demand high-price parts, most stockouts are satisfied by reactive lateral or emergency shipments. Yet, 30 out of 382 stockouts (7.85% of total stockouts) are resolved by backordering and waiting for pipeline stock. Therefore, despite the high downtime penalty charge, waiting for pipeline stock can be selected as a preferred solution.

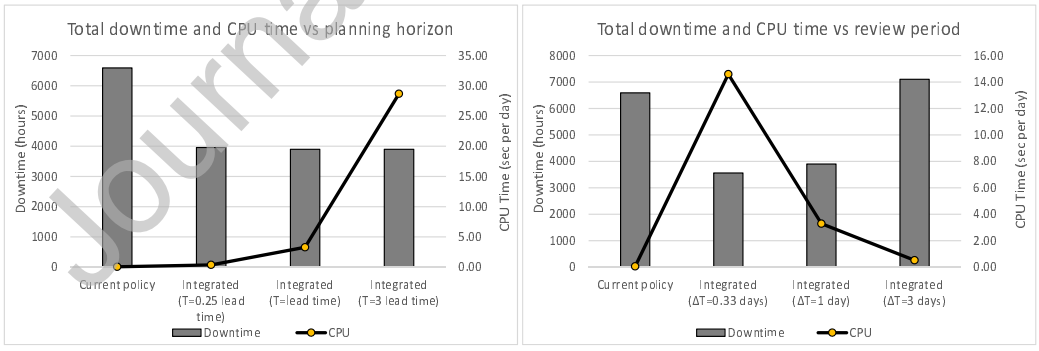
### 5.3. Impact of parameters and complete pooling at the downstream on the value of our approach

We investigate the impact of downtime penalty charge, planning horizon, and review period. Keeping other parameter values same, we consider (i) 10, 50, 2500 (which is the base case value) and 75000 Euros for downtime penalty charge, (ii) 0.25, 1 and 3 times the replenishment lead time for each part for planning horizon, (iii) $\frac{1}{3}$, 1, and 3 days for review period. We investigate the impact of proactive interventions which are slower but also cheaper than their reactive counterparts. In these experiments, we take proactive lead times as 2 times the reactive lead times and shipment costs of proactive

interventions as $\frac{1}{2}$ of shipment costs of reactive interventions. This is as opposed to assuming identical values for both intervention types. We also test our approach with different base stock levels. We generate base stock levels by selecting arbitrary individual fill rates and finding minimum stock levels that guarantee these fill rates. Based on the results, we make the following observations:

- *The performance of our integrated approach deteriorates with shorter planning horizons (slightly) and longer review periods:* Figure 5a illustrates the impact of the planning horizon length for the considered values. The figure reveals that downtime is (slightly) higher when planning horizon is shorter. Setting planning horizon to regular replenishment lead time performs well. The average CPU time (per day) increases with the planning horizon. Figure 5b shows the impact of the review period for $\frac{1}{3}$, 1 and 3 days. The figure reveals that total downtime increases with the review period. Increasing the review period makes proactive lateral transshipment and proactive emergency shipments less attractive since they contribute less to lead time and downtime reduction. Figure 5b also shows that the average CPU time decreases with the review period.

Figure 5: Effect of planning horizon and review period on total downtime and average CPU time.



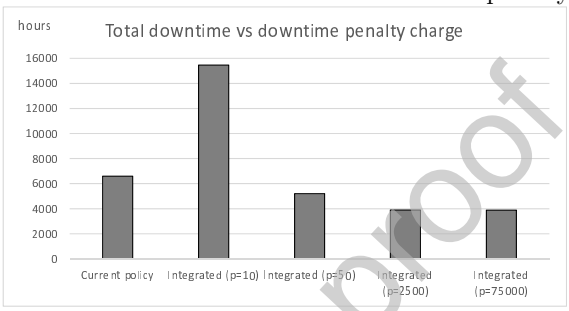

a) Effect of planning horizon on total downtime and average CPU time

b) Effect of review period length on total downtime and average CPU time

- *Only low downtime penalty charge has a significant impact on downtime:* Figure 6 demonstrates the total downtime for penalty charge of 10, 50,

2500 and 75000 Euros. As seen in this figure, the performance of our integrated approach is insensitive to downtime penalty charge when downtime penalty charge is e2500 or higher. On the other hand, it may perform arbitrarily badly for very low values of downtime penalty cost. This is due to low service level caused by low penalty cost.

Figure 6: Total downtime under different downtime penalty charges.



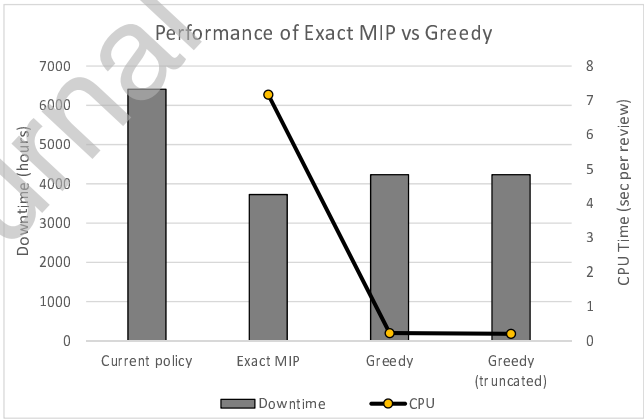

- *Allowing complete pooling reduces the downtime further:* We test the value of complete pooling by allowing reactive and proactive lateral transshipments between all warehouses in the global supply chain. The total downtime is 3536 compared to 3898 when the integrated approach is applied to current network. Also, with complete pooling, the proportion of proactive lateral transshipments and the proportion of proactive emergency shipments to total proactive interventions increase substantially to 39% and 51%, respectively.

- *Downtime reduction is sensitive to tactical base stock levels, shipment lead times for proactive interventions and less sensitive to shipment costs of proactive interventions:* We find that the effectiveness of (near-) optimal operational decisions depend on the base stock levels. When base stock levels are high, the improvement potential of using operational planning reduces, and vice versa. Since tactical planning is outside our scope, we do not address this issue further. We also find that downtime remains at 3949 hours when proactive shipment costs decrease by half whereas downtime increases to 4078 hours when shipment lead times for proactive interventions double.

### 5.4. Performance of the greedy heuristic and the impact of truncation

Figure 7 summarizes the performance of the greedy heuristic against the exact MIP solution. The gap between the greedy heuristic and the exact MIP solution is relatively large. This contradicts to a certain extent the observation that the greedy heuristic performs substantially well in spare parts applications (Kranenburg and van Houtum, 2009, and Topan et al., 2017). This difference is attributed to our multi-period setting and the characteristics of the greedy heuristic. The greedy heuristic, being myopic in nature, does not account for orders which can be placed later during the planning horizon (see equations (6)-(10)). In contrast to the greedy heuristic, the exact MIP formulation take these future orders into account through constraints (2)-(4). Yet, the greedy algorithm makes significant improvements over the current policy. Furthermore, although we do not show results here, in several numerical results we observe that the gap between the exact MIP and the greedy heuristic vanishes when base stock levels are higher and the likelihood of stockout at the depot is zero. Finally, Figure 7 (see *Greedy truncated*) shows that the truncation of the greedy heuristic has almost no impact on the solution quality of the greedy heuristic.

Figure 7: The performance of the greedy heuristic against the exact MIP and the impact of truncation.




### 5.5. Computational e ciency and experiments with very large instances

We test our model and its exact solution with a larger instance. We select 360 parts by ranking parts according to an ABC classification. The

selected parts represent approximately 90% of the total turnover). We use the same experimental setting as in our previous experiments. The average computation time is half a second per item per period per replication. This includes both reactive and proactive intervention decisions by solving the MIP exactly for a finite horizon $T$. Therefore, the computational efficiency of our method is quite appealing. Using our integrated approach, the total downtime reduces by one-third (slightly more than 30000 hours). As opposed to the downtime reduction, the total shipment cost increases by 2.5 million Euros. Note that this additional cost is relatively small compared to the gains from downtime reduction.

## 6. Conclusion and open issues

In this paper, we consider operational level planning of spare parts in a multi-item two-echelon inventory system. We propose a generic model that integrates decisions on reactive and proactive interventions. To test our model, we conduct a numerical experiment using the data of a leading manufacturer, which operates an extensive supply network for its spare parts.

Based on the numerical study, we make the following main observations:

- The integrated approach reduces downtime substantially, and in this way, leads to considerable cost savings. It is also computationally efficient enough to solve large scale problems. In that respect, our findings are quite promising for future use of our model in practice.
- The simple approach for reactive intervention decisions, satisfying the demand from the nearest location with positive stock, performs very well. This indicates that the downtime reduction stems mainly from optimal proactive interventions. Yet, we also find that leaving out reactive interventions and solving stockouts at periodic review points leads to high downtime.
- Among two proactive intervention types, proactive emergency shipments contribute most to downtime reduction. For low price parts, proactive lateral transshipments also have a significant contribution to downtime cost reduction. The benefit of our integrated approach is higher for high demand parts.
- Allowing complete pooling increases downtime savings and usage of both proactive intervention types further.

Our findings rely on the data of the manufacturer, which includes spare parts that are predominately slow moving and has limited part pooling between warehouses due to customs regulations. Also, as proactive lateral transshipments are currently not used by the manufacturer, costs of proactive interventions and their reactive counterparts are assumed to be equal. This overestimates the costs of proactive interventions, which are typically not as urgent as their reactive counterparts. Therefore, the value of our integrated approach could be higher for a setting with higher demand rates, complete pooling, and less expensive proactive interventions.

We identify the following open issues: First, we follow a hierarchical approach and assume that the feedback between tactical planning and operational planning is one-way. However, the effectiveness of (near-) optimal planning operational decisions depend on how good the tactical planning and operational planning are coordinated. Joint optimization or coordination of operational and tactical level planning requires further attention. Second, the service level agreements committed to customers set either explicit or implicit short-term operational service targets usually defined for each operational planning horizon. Yet, customer differentiation has mainly been addressed at tactical level in the literature, and the studies about operational level customer differentiation are limited. This raises questions such as how to operate a spare parts network under operational level service targets and how to pool common parts in the supply network exploiting the differences in operational level customer service targets and real-time service levels. These questions need to be addressed. Third, we assume that supplier lead times are fixed. However, in practice supplier lead times are uncertain. This has an adverse impact on the overall performance of the supply chain. We see in practice that supplier lead times are quite variable. Real-time supply information, e.g., expected delivery times, could be exploited to counter negative effects of supply uncertainty. This is a topic for further research.



## Acknowledgment

This research is part of the project on Proactive Service Logistics for Advanced Capital Goods Next and has been sponsored by TKI-Dinalog (Dutch Institute for Advanced Logistics). The authors would like to thank Martijn and Ruud for the discussions that formed the basis of this study.

## References


Arts, J., Basten, R., and van Houtum, G. J. (2016). "Repairable stocking and expediting in a fluctuating demand environment: Optimal policy and heuristics", *Operations Research*, Vol. 64(6), 1285-1301.

van Aspert, M., (2015). "Design of an integrated global warehouse and field stock planning concept for spare parts", *PDEng thesis*, Technische Universiteit Eindhoven, Eindhoven.

Basten, R.J.I. and van Houtum, G.J., (2014). "System-oriented inventory models for spare parts", *Surveys in Operations Research and Management Science*, Vol. 19, 34-55.

Caggiano, K. E. , Muckstadt, J.A., and Rappold, J.A., (2006). Integrated Real-Time Capacity and Inventory Allocation for Reparable Service Parts in a Two-Echelon Supply System", *Manufacturing and Service Operations Management*, Vol. 8 (3), 292-319.

Grahovac, J., and Chakravarty, A., (2001). "Sharing and Lateral Transshipment of Inventory in a Supply Chain with Expensive Low-Demand Items", *Management Science*, Vol. 47 (4), 579-594.

Glazebrook, K., Paterson, C., Rauscher, S., and Archibald T., (2015). "Benefits of Hybrid Lateral Transshipments in Multi-Item Inventory Systems under Periodic Replenishment", *Production and Operations Management*, Vol. 24 (2), 311-324.

van der Heijden, M. C., Diks, E. B., and de Kok, A. G., (1997). "Stock allocation in general multi-echelon distribution systems with 4R1 S5 order-upto- policies", *Internatioanl Journal of Production Economics*, Vol. 49(2) 157174.

Hoadley, H., and Heyman, D.P., (1977). "A two-echelon inventory model with purchases, dispositions, shipments, returns and transshipments", *Naval Research Logistics*, Vol. 24, 1-19.

van Houtum, G.J., and Kranenburg, B., (2015). *Spare Parts Inventory Control under System Availability Constraints*. Springer.

Howard, C, Marklund, J., Tan, T., and Reijnen, I., (2015). "Inventory Control in a Spare Parts Distribution System with Emergency Stocks and Pipeline Information", *Manufacturing and Service Operations Management*, Vol. 17 (2), 142-156.

Kranenburg, A.A., and van Houtum, G.J., (2009). "A new partial pooling structure for spare parts networks", *European Journal of Operational Research*, Vol. 199, 908-921.

Marklund, J., Rosling, K., (2012). "Lower Bounds and Heuristics for Supply Chain Stock Allocation", *Operations Research*, Vol. Vol. 60 (1), 92-105.

Muckstadt J. A., (2005). *Analysis and Algorithms for Service Parts Supply Chains*, Springer, New York.

Paterson, C., Teunter, R., and Glazebrook, K., (2012). "Enhanced lateral transshipments in a multi-location inventory system", *European Journal of Operational Research*, Vol. 221, 317-327.

Paterson, C., Kiesmller, G., Teunter, R., and Glazebrook, K. (2011). "Inventory models with lateral transshipments: A review", *European Journal of Operational Research*, Vol, 210 (2), 125-136.

Silver, E.A., Pyke, D., and Peterson, R., (1998). Inventory Management and Production Planning and Scheduling, 3rd

Sherbrooke, C.C., (2004). *Optimal Inventory Modeling of Systems. Multi-Echelon Techniques*, Kluwer, Dordrecht, The Netherlands.

Song, J.S., Zipkin P (2009). "Inventories with multiple supply sources and networks of queues with overflow bypasses", *Management Science*, Vol. 55 (3), 362-372.

Tiemessen, H.G.H., Fleischmann, M., van Houtum, G.J., van Nunen, J.A.E.E., and Pratsini, E., (2013). "Dynamic demand fulfillment in spare parts networks with multiple customer classes", *European Journal of Operational Research*, Vol. 228, 367-380.

Thonemann, U.W., (2002). "Improving supply-chain performance by sharing advance demand information", *European Journal of Operational Research*, Vol. 142 (1), 81-107.

Topan, E., Bayındır, Z.P., and Tan, T., (2017). "Heuristics for multi-item two-echelon spare parts inventory control subject to aggregate and individual service measures", *European Journal of Operational Research*, Vol. 256, 126-138.

Topan, E., Eruguz, A.S., Ma, W., van der Heijden, M., and Dekker, R., (2019). "A review of operational spare parts service logistics in service control towers", *European Journal of Operational Research*.

Veeraraghavan S, Scheller-Wolf, A., (2008). "Now or later: A simple policy for effective dual sourcing in capacitated systems", *Operations Research*, Vol. 56 (4), 850-864.