Manufacturing & Service Operations Management

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Strategic Safety Stock Placement in Supply Networks with Static Dual Supply

Steffen T. Klosterhalfen, Stefan Minner, Sean P. Willems

MANUFACTURING & SERVICE OPERATIONS MANAGEMENT
Vol. 16, No. 2, Spring 2014, pp. 204–219
ISSN 1523-4614 (print) | ISSN 1526-5498 (online)

http://dx.doi.org/10.1287/msom.2013.0472

Strategic Safety Stock Placement in Supply Networks with Static Dual Supply

Steffen T. Klosterhalfen
BASF SE, GVM/S, 67056 Ludwigshafen, Germany, steffen.klosterhalfen@basf.com

Stefan Minner
TUM School of Management, Technische Universität München, 80333 Munich, Germany, stefan.minner@tum.de

Sean P. Willems
School of Management, Boston University, Boston, Massachusetts 02215, willems@bu.edu

Many real-world supply networks source required materials from multiple suppliers. Existing multiechelon inventory optimization approaches either restrict their scope to multiple supply sources in two-echelon systems or single suppliers in multiechelon systems. We develop an exact mathematical model for static dual supply in a general acyclic N-echelon network structure, which builds on the guaranteed-service framework for safety stock optimization. It is assumed that the suppliers are allocated static fractions of demand. We prove that for normally distributed demand an extreme point property holds. We present a real example from the industrial electronics industry consisting of five echelons and three dual-sourced materials. This example forms the basis for a numerical analysis. Compared with the only previously published approximate solution, our exact approach results in considerable cost savings because the exact model captures inventory pooling in a way that the approximation is unable to do. For a set of test problems, total safety stock cost savings are 9.1%, on average.

Keywords: dual sourcing; guaranteed service; multiechelon inventory system; safety stock optimization
History: Received: July 15, 2012; accepted: October 10, 2013. Published online in *Articles in Advance* March 7, 2014.

1. Introduction

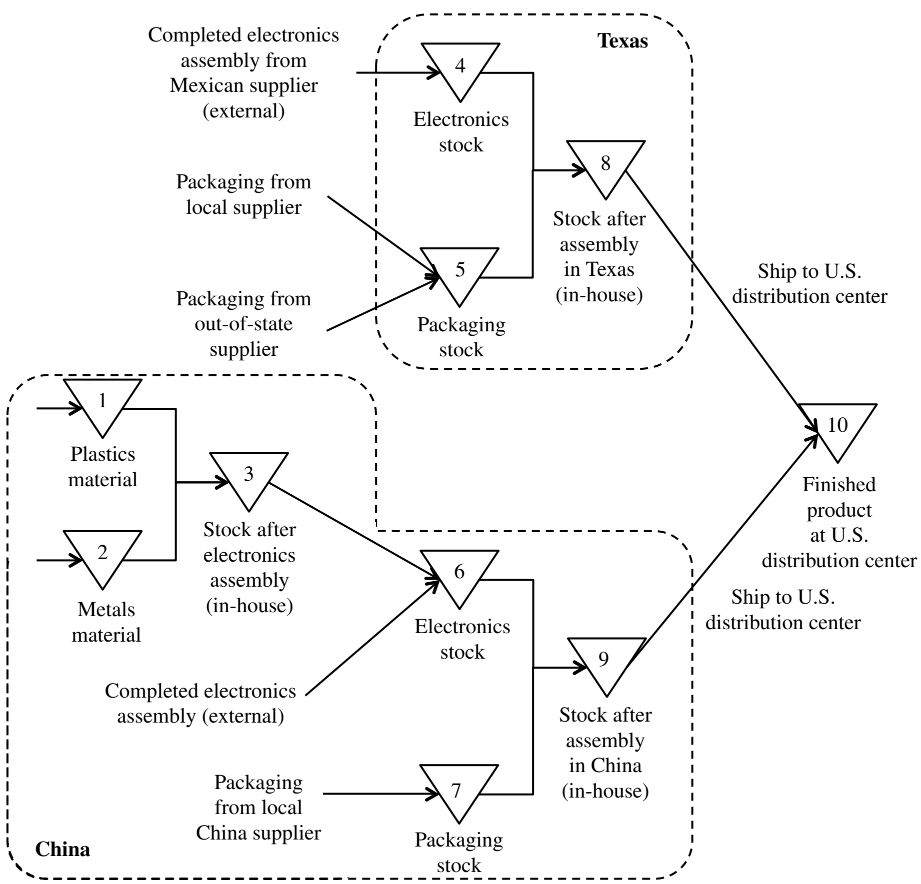
The last two decades have witnessed the emergence of strategic safety stock optimization models. These models are strategic in the sense that they optimize both safety stock levels and safety stock holding locations across the multiechelon supply chain. To maintain tractability, these models focus less on the operational facets of inventory management at any one location and instead consider the safety stock impacts on time, variability, and cost across the supply chain. Integrating multisourcing with inventory optimization is a logical progression of this work.

Existing multisourcing research has focused on operational and tactical issues like the optimal dynamic replenishment policy in the presence of multiple suppliers. However, consistent with strategic safety stock optimization, multisourcing also has important strategic dimensions we have observed in practice. First, multisourcing mitigates risk relative to a single supply source. Second, multisourcing increases the likelihood of securing upside volume since multiple suppliers are capable of providing the item. Although one can also envision negotiating leverage and cost containment as reasons for multisourcing, in the companies we have worked with, risk mitigation and supply assurance were the more significant motivators for multisourcing. Furthermore, at the supply chain design phase where we have witnessed this problem, the sourcing teams are not trying to precisely define upside, supply assurance, or risk mitigation. Instead, they are trying to find a pragmatic inventory stocking solution across a complex supply chain network. Consistent with this strategic perspective on multisourcing, the relative fraction of demand allocated to each supplier is primarily a function of supplier capability and capacity versus lowest cost; in cases where one supplier is significantly more expensive than another supplier, the more expensive supplier is simply allocated the lowest amount possible to maintain a feasible multisourcing relationship. Namely, Intel Corporation has identified a minimum of 20% demand volume allocated to sources that supply an item they have decided must be multisourced (Tian et al. 2011). In cases where the cost disparity is not great, the volume allocated to each supplier is roughly equal.

To reflect this reality in our model, we adopt an order-splitting policy among the suppliers. At an operational and tactical level, which most contributions in the dual-sourcing literature focus on, an order-splitting policy usually optimizes the supply fractions and thus order quantities with the suppliers dynamically over time. At the strategic level we are considering, we assume that the order split is static across time. In the two-supplier case, it means that the demand in period t at node i, $d_{i,t}$, is split into order quantities $\delta_{j,i} \cdot d_{i,t}$ and $(1-\delta_{j,i}) \cdot d_{i,t}$, where $\delta_{j,i}$ denotes the supply fraction of node i with supplier j, which is the same for every period t. To rationalize the assumption of a static supply fraction, it is useful to draw an analogy to hierarchical production planning. Assuming a static supply fraction at the strategic planning level allows for the solution of the safety stock planning problem across the entire network. This does not imply that, in the short run, order quantities from suppliers never deviate from the static allocations. It simply implies that the scope of the strategic problem does not, and cannot, encompass these short-term deviations. The approach taken is similar to hierarchical production planning, where detailed scheduling aspects are neglected when making long-term planning decisions. Boute and van Mieghem (2013) resort to a similar static dual-sourcing policy even in a single-echelon setting to make the problem analytically tractable and derive an exact formula for the optimal value of what they term the "strategic allocation." The technical advantage of the assumption of static supply fractions also holds true in our multiechelon setting because it allows us to characterize upstream demand independent from downstream safety stock decisions, i.e., external demands propagate upstream according to the splitting ratio.

Figure 1 depicts a supply network where two upstream nodes (5 and 6) and the final node (10) have two suppliers for the same material. Each node, represented by a triangle, denotes a possible holding location for safety stock. Arcs denote processing activities between nodes. The illustrated product belongs to a large multinational corporation that focuses on industrial electronics products for the construction industry. The essence of this product is an electrical component and its packaging. The electronic assembly is made at three sites. One site is internal (in China), which satisfies about 50% of the total demand, and the other two are outsourced suppliers. For the internal site, metal and plastic raw materials are converted to the electronics assembly (nodes 1, 2, and 3). For the two external sites, the unit of measure is the completed electronics assembly (nodes 4 and 6). Final assembly occurs at one of two internal manufacturing sites. One is located in Texas (node 8), and the

Figure 1 **Supply Network with Five Echelons and Three Dual-Sourced Materials**



other one in China (at the same location that makes the electronic assembly, node 9). For the Texas site, packaging comes from one of two packaging suppliers (node 5); 40% comes from a local supplier in Texas and the other 60% comes from an out-of-state supplier. In China, all the packaging comes from one supplier (node 7). After the final assembly, the finished product is shipped to a distribution center in the central United States (node 10). This is accomplished by truck from the Texas facility and by a combination of sea and truck for the China facility.

The problem in Figure 1 is common in practice. For example, Intel employs multiple supply options at several echelons of its chip-making supply chain, including multiple raw silicate suppliers, wafer fabs, assembly and test facilities, and packaging operations. ExxonMobil Chemicals employs multiple feedstock suppliers and chemical intermediates plants that supply several plasticizer operations making finished goods, including flexible piping and cable sheathing.

In this paper, we formulate a multiechelon inventory model, suitable for strategic planning, that takes static dual supply into account at different locations in a general acyclic supply network. The determination of optimal safety stock levels and locations in multiechelon models already represents a challenging task. Incorporating multiple supply sources into such settings increases the complexity even further. To make progress on these more complex systems, we sacrifice some of the rigor found in the literature on multiechelon systems for the benefit of capturing sourcing aspects that are relevant to practice and computationally manageable. Therefore, we place our multiechelon dual-supply model within the guaranteed-service (GS) framework, the practical relevance of which is demonstrated by its widespread implementation at companies such as Hewlett-Packard (Billington et al. 2004), Intel (Manary and Willems 2008), and Procter & Gamble (Farasyn et al. 2011). For the safety stock computation under the GS assumption, we derive exact expressions for the demand variability over the coverage time in the presence of two suppliers allocated static fractions of demand. To obtain this exact characterization, we show that we need to switch from an activity-on-node network formulation, which is employed in all previous GS papers, to an activity-on-arc formulation. By comparing our exact approach with the only previously published approximate one in Graves and Willems (2005), we demonstrate how the exact approach accurately captures inventory pooling at a dual-sourced node, which the approximate one is unable to do. This may result in different safety stock positioning within the network or considerably less stock at the same nodes. We prove that for normally distributed demand an extreme point property holds for the exact optimization problem. This limits the set of optimal solution candidates, which can be useful in the development of solution algorithms.

The remainder of this paper is organized as follows. Section 2 presents a literature review. The notation and assumptions are outlined in §3. Section 4 develops the mathematical model, first for a single node and then for general acyclic networks. The developed approach is applied to the real-world example in §5, which also forms a test suite to compare this approach with an approximate one. Section 6 provides a summary. All proofs are relegated to the online appendix (available as supplemental material at http://dx.doi.org/10.1287/msom.2013.0472).

2. Literature

For periodically reviewed supply networks with a single source of supply at each node, various contributions exist that address the problem of safety stock optimization. Following the nomenclature in Graves and Willems (2003), a broad classification can be made into stochastic-service (SS) and GS models based on the pioneering works by Clark and Scarf (1960) and Simpson (1958), respectively. Because we use the GS approach as a model basis, we refer the reader to van Houtum (2006) for details about SS contributions. In contrast to the SS approach, which assumes that safety stock is the only available buffer against demand fluctuations, the GS framework assumes that further countermeasures, beyond safety stock, are available to satisfy demand. Thus, no stochastic delay in the delivery of ordered items is propagated in the system. Over the last two decades, the GS approach has been extended in several ways. Extensions include optimizing safety stock placement in different network structures including even general acyclic ones (see Humair and Willems 2011 and references therein), under nonstationary demand (Graves and Willems 2008), and under evolving forecasts (Schoenmeyr and Graves 2009), to name a few.

For supply networks with multiple suppliers, in particular dual sourcing, the existing inventory literature primarily deals with the single-echelon context. Minner (2003) provides a comprehensive overview of the existing models up to around 2001. Whereas early contributions mainly focus on finding the structure of the optimal policy for special cases (Daniel 1962, Fukuda 1964, Whittemore and Saunders 1977), later works address the development of heuristic policies for more general settings and their parameter optimization. Prominent examples for the periodic-review case include the standing-order policy (Rosenshine and Obee 1976, Chiang 2007), the constant-order policy (Janssen and de Kok 1999), and the dual-index policy (Veeraraghavan and Scheller-Wolf 2008, Arts et al.

2011), among others. Recently, because of the established relationship between the dual-sourcing and the lost-sales inventory problem by Sheopuri et al. (2010), policies that show a good performance for the latter problem are transferred to the dual-sourcing problem. These include policies that resemble an order-splitting policy (see an overview by Thomas and Tyworth 2006), where the splitting decision, which is usually made once for all periods, is adjusted each period.

Multiechelon models with two or more supply sources are limited to two echelons in the form of one-warehouse multiple-retailer systems operating under continuous review. Ganeshan (1999) presents a near-optimal (s, Q) policy, when the warehouse order is split across several identical suppliers. Various other works consider low-demand items; see, e.g., Muckstadt and Thomas (1980) and Aggarwal and Moinzadeh (1994). Another related multiechelon research stream concerns models with order expediting; see, e.g., Lawson and Porteus (2000) or Berling and Martínez-de-Albéniz (2012). Those models consider different modes of supply for an item in terms of processing speed.

The objective of our paper differs from the existing single-echelon and multiechelon models with multiple suppliers in two respects. First, we are not trying to model an optimal replenishment policy at an operational level of detail. We rather take a more strategic view and focus on the placement of safety stock across the network. Based on the empirical evidence provided in the introduction and for reasons of analytical tractability, we assume that a static allocation of the demand shall be assigned to each supplier in every period at a strategic planning level. The static allocation is treated as exogenous based on long-run time averages and company-specific policies, such as minimum production quantities for facilities. (We elaborate on this assumption in §3.3 when explaining the replenishment policy.) Second, we consider larger networks with multiple suppliers and multiple echelons. To this end, we build on the GS approach for multiechelon inventory optimization, which has been shown to be applicable to supply networks of larger sizes (see Willems 2008).

So far, all GS contributions assume that each item is sourced from a single supplier. Solely, Graves and Willems (2005) briefly outline an approximate approach for how to incorporate multiple suppliers in the final section of their paper. We derive exact expressions, and in §5, we show that this exact approach outperforms the approximate one.

3. Notation and Assumptions

We use the GS framework as the basis for our model, summarizing its assumptions in this section.

3.1. Supply Network Modeling

The supply chain is modeled as a network of n nodes connected by arcs. In the existing GS literature, nodes are stages where each stage represents a major processing function and a potential location to hold safety stock after the process has finished. Arcs simply denote that an upstream stage supplies a downstream stage. Unlike this activity-on-node network representation, to model multisourcing, we need to adopt an activity-on-arc network representation. In the activity-on-arc network representation, nodes are exclusively potential locations to hold safety stock, and all processing activity takes place on the arcs. Section 5.1 contains an example illustrating the difference between the two approaches.

Let $\mathcal{A}$ be the arc set for the network representation, and let $(i, j) \in \mathcal{A}$ denote the arc between nodes i and j, indicating that the items are processed as they move from node i to j. With each arc, we associate a scalar $a_{i,j}$ (also called the production coefficient) that indicates how many items of upstream node i are required per unit at node j. External suppliers with ample stock have a node index 0 (or a negative one to distinguish between them when necessary). All nodes supplying external customers are summarized in set E. Out of these nodes, the ones requiring items from node i are collected in set $E(i)$.

3.2. Processing Times

Before entering a node, an item passes through a process, e.g., a manufacturing or transportation step, represented by an arc. We assume a node can source a required item from up to two suppliers. The processing time between nodes i and j is denoted by $T_{i,j}$, which is deterministic and known. There are no capacity constraints on the production or transportation between nodes.

The distinction between a (static) dual-supply situation and an assembly situation is made by the use of different arc types. Whereas Figure 2 refers to the dual-supply case with $T_{1,3}$ likely having a different value than $T_{2,3}$, Figure 3 illustrates an assembly situation where nodes 1 and 2 provide different items for the assembly process preceding node 3, which takes $T_{1,3} = T_{2,3}$ periods.

Figure 2 Static Dual-Supply Situation

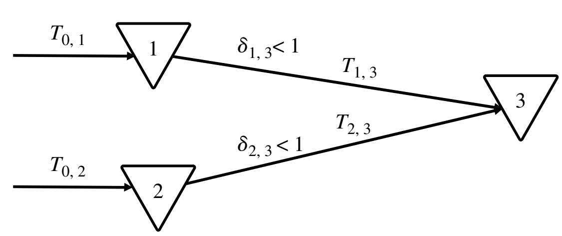
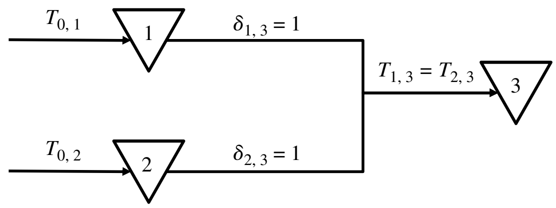


Figure 3 **Assembly Situation**



3.3. Replenishment Policy

In line with the standard GS approach, we assume that the replenishment policy at each node follows standard base-stock control (with a common review period), where in each period node i places an order that raises its inventory position to the desired base-stock level, B_i. However, under the dual-supply regime, only a fraction of the entire demand, denoted $\delta_{j,i}$, is placed with each supplier j. (Obviously, for a node with only a single supplier, $\delta_{j,i} = 1$.) We assume that this fraction is static; i.e., it is not adapted from one period to the other. Moreover, we assume the supply fraction is exogenously determined and not an optimization model output. We do this for two reasons. First, Klosterhalfen (2010) presents an optimization model that treats the static supply fractions as decision variables. However, the formulation is significantly more complex, and given this complexity, the solution does not lend itself to interesting insights. As such, this paper focuses on what we consider to be the innovative facets of this research. Second, in Klosterhalfen (2010) as well as in most of the single-echelon dual-sourcing literature, only the trade-off between procurement cost and inventory holding cost is analyzed to determine an optimal supply fraction. At the strategic level we are considering, the sourcing teams are taking a higher-level view of sourcing fractions. As noted in Tian et al. (2011), if two suppliers have radically different costs, then Intel maintains a minimum allocation to the expensive supplier. The company represented in Figure 1 of this paper follows a similar policy, and when costs are not radically different, it tries to allocate the same volume to each supplier with the idea that this can help secure upside volume in the future. This approach is also consistent with Allon and van Mieghem (2010), who analyzed a more complex sourcing policy but found a ratio of 75:25 as a reasonable rule for allocating demand between slow and fast supply sources on a strategic level.

As a result of the base-stock replenishment policy with static supply fractions, the replenishment order placed by node i with supplier j in period t is $Q_{(j,i),t} = \delta_{j,i} \cdot a_{j,i} \cdot d_{i,t}$, where $d_{i,t}$ denotes the demand in period t at node i. Apart from the arc types (see Figures 2 and 3), the supply fractions $\delta_{j,i}$ are also an indicator for dual or single supply, i.e., $\delta_{i,j} \in (0,1) \Rightarrow$ dual supply and $\delta_{i,j} = 1 \Rightarrow$ single supply.

3.4. Demand Process

We assume the following sequence of events in each period: outstanding orders arrive at each node at the beginning of the period, external demand for that period occurs and is propagated through the network also at the beginning of the period as a result of the specified replenishment policy, external and internal demand is satisfied from each node, and costs are assessed at the end of the period.

According to the replenishment policy, all nodes see the external customer demand in a period adjusted by the supply fractions and production coefficients. Demand at an internal node i is

$$d_{i,t} = \sum_{(i,j)\in\mathcal{A}} \delta_{i,j} \cdot a_{i,j} \cdot d_{j,t} \quad \forall\, i \notin E. \tag{1}$$

External demand per period at node $e \in E$, D_e, is assumed to follow a normal distribution with stationary mean μ_e and standard deviation σ_e and is identically and independently distributed over time. For items at demand nodes e and f in an arbitrary period, $\rho_{e,f}$ denotes the correlation between demands. Then, for node $i \notin E$, we get

$$\mu_i = \sum_{e\in E(i)} \delta_{i,e} \cdot a_{i,e} \cdot \mu_e, \tag{2}$$

$$\sigma_i^2 = \sum_{e\in E(i)} \sum_{f\in E(i)} \delta_{i,e} \cdot a_{i,e} \cdot \delta_{i,f} \cdot a_{i,f} \cdot \sigma_e \cdot \sigma_f \cdot \rho_{e,f}, \tag{3}$$

with $a_{e,e} = \delta_{e,e} = \rho_{e,e} = 1$, $\forall\, e \in E$, and the formulas below (recursively applied) for all $i \notin E$:

$$a_{i,e} = \sum_{(i,j)\in\mathcal{A}} a_{i,j} \cdot a_{j,e} \quad \forall\, i \notin E, \tag{4}$$

$$\delta_{i,e} = \sum_{(i,j)\in\mathcal{A}} \delta_{i,j} \cdot \delta_{j,e} \quad \forall\, i \notin E. \tag{5}$$

Let $D_i(L)$ denote the L-period demand random variable at node i.

3.5. Guaranteed Service Times

The outgoing service time for node i, S_i, is the amount of time that elapses between the placement of an order by a downstream node j, such that $(i,j) \in \mathcal{A}$, and the fulfillment of this order by node i; i.e., any material request (irrespective of the quantity) by node j made at time t is available for shipment or processing after S_i periods, and this order will enter into inventory at node j at time $t + S_i + T_{i,j}$. This in fact means that the outgoing service of a node does not depend on the current physical material availability, but it is assumed that necessary countermeasures will be adopted to make missing items available in time. For more details about this guaranteed-service assumption, see, e.g., Graves and Willems (2003). Service times are the decision variables for the optimization model, as will be seen in §4. As in the standard GS literature, we assume that a node quotes the same outgoing service time to all downstream nodes.

4. Mathematical Model Formulation

4.1. Single-Node Analysis

4.1.1. Inventory Model. Consider a setting where node i has two suppliers j and k (Figure 4). Given the GS assumptions, the items of node j with $(j, i) \in \mathcal{A}$ are available for processing after S_j periods. Then, it takes another $T_{j,i}$ periods until the items are in stock at i. Consequently, each replenishment placed with supplier j is received after a replenishment lead time $S_j + T_{j,i}$. The same holds for supplier k. According to the replenishment lead times, we can identify the slow and fast supplier of node i. We classify all corresponding variables and parameters by assigning a superscript s for slow. Thus, we have S_i^s for the outgoing service time of the slow supplier to node i, T_i^s for the processing time between the slow supplier and node i, δ_i^s for the supply fraction, a_i^s for the production coefficient, and $Q_{i,t}^s$ for the order quantity in period t. Analogously, the variables and parameters of the fast supplier to i receive superscript f.

Node i itself quotes an outgoing service time S_i to any downstream adjacent nodes. As a consequence, three potential replenishment-lead-time service-time cases may occur at node i:

1. The outgoing service time S_i is shorter than (or equal to) both replenishment lead times, i.e., $S_i^s + T_i^s \geq S_i^f + T_i^f \geq S_i$.
2. The outgoing service time S_i is larger than (or equal to) the fast replenishment lead time but shorter than (or equal to) the slow replenishment lead time, i.e., $S_i^s + T_i^s \geq S_i \geq S_i^f + T_i^f$.
3. The outgoing service time S_i is larger than (or equal to) both replenishment lead times, i.e., $S_i \geq S_i^s + T_i^s \geq S_i^f + T_i^f$.

We introduce SI_i^s (SI_i^f) as the incoming service time from the slow (fast) supplier to node i. Just as all nodes quote an outgoing service time to any downstream-adjacent node(s), each node is quoted an incoming service time from any upstream-adjacent node(s). From the perspective of the upstream-adjacent node, this is its outgoing service time. From the perspective of the downstream-adjacent node, this is the incoming service time quoted to the downstream-adjacent node from the upstream-adjacent node. As in the standard GS literature (see, e.g., Graves and Willems 2000), we permit $SI_i^s \geq S_i^s$ ($SI_i^f \geq S_i^f$) to allow for the possibility that the replenishment lead time for node i, $S_i^s + T_i^s$ ($S_i^f + T_i^f$), is less than its outgoing service time S_i. In this case, items would arrive at node i before they are actually needed for filling i's demand. To avoid unnecessary inventory, node i would ideally inform the respective supplier to delay its delivery, i.e., increase its service time. The permitted delay from each supplier (in addition to its current outgoing service time) is chosen such that the effective replenishment lead time for node i equals its outgoing service time S_i. Therefore, we define the incoming service times from the slow and fast supplier as

$$SI_i^s = \max\{S_i - T_i^s, S_i^s\} \quad \text{and} \qquad SI_i^f = \max\{S_i - T_i^f, S_i^f\}. \tag{6}$$

From (6), it is obvious that the incoming service times at i can be directly derived from the suppliers' outgoing service times, which are our decision variables, and vice versa. We find the following:

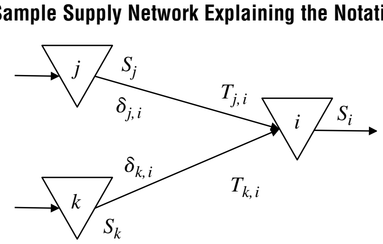
Proposition 1. *All replenishment-lead-time service-time cases* (i) $S_i^s + T_i^s \geq S_i^f + T_i^f \geq S_i$, (ii) $S_i^s + T_i^s \geq S_i \geq S_i^f + T_i^f$, *and* (iii) $S_i \geq S_i^s + T_i^s \geq S_i^f + T_i^f$ *reduce to the following relation*:

$$SI_i^s + T_i^s \geq SI_i^f + T_i^f \geq S_i. \tag{7}$$

As a result of Proposition 1, we focus on relation $SI_i^s + T_i^s \geq SI_i^f + T_i^f \geq S_i$ in the derivation of the net stock expression and the upcoming exposition. Without loss of generality, we assume that the net stock at the end of period 0 is equal to the base-stock level: $IL_{i,0} = B_i$. Then, the net stock at the end of period t is

$$IL_{i,t} = B_i - \underbrace{\sum_{m=1}^{t} d_{i,m}}_{(a)} + \underbrace{\sum_{m=1}^{t-(SI_i^s+T_i^s)} \frac{Q_{i,m}^s}{a_i^s}}_{(b)} + \underbrace{\sum_{m=1}^{t-(SI_i^f+T_i^f)} \frac{Q_{i,m}^f}{a_i^f}}_{(c)} + \underbrace{\sum_{m=t-(S_i-1)}^{t} d_{i,m}}_{(d)}$$

Figure 4 Sample Supply Network Explaining the Notation



$$= B_i - \sum_{m=1}^{t-S_i} d_{i,m} + \delta_i^s \cdot \sum_{m=1}^{t-(SI_i^s+T_i^s)} d_{i,m} + \delta_i^f \cdot \sum_{m=1}^{t-(SI_i^f+T_i^f)} d_{i,m}$$

$$= B_i - \underbrace{\left(\sum_{m=t-(SI_i^f+T_i^f-1)}^{t-S_i} d_{i,m} + \delta_i^s \cdot \sum_{m=t-(SI_i^s+T_i^s-1)}^{t-(SI_i^f+T_i^f)} d_{i,m} \right)}_{(e)}, \tag{8}$$

where $(b)+(c)$ represents the received replenishments up to period t, and $(a)-(d)$ represents the realized demand that has been delivered to the successor(s) up to period t. After some manipulations, the random variable that characterizes the demand that needs to be covered by safety stock is given by (e). We call it the coverage random variable.

Figure 5 illustrates the timeline for the calculation. First, in case of a positive outgoing service time S_i, the fulfillment of the demand in period t can be delayed by S_i periods. Moreover, the orders placed with the slow supplier up to and including period $t-(SI_i^s+T_i^s)$ have arrived in period t, i.e., the δ_i^s-fraction of those period demands. Only the remaining orders up to period t are still outstanding. Out of those, only the orders up to $t-S_i$ are relevant for the coverage random variable. Similarly, the outstanding orders with the fast supplier are given by the $(1-\delta_i^f)$-fraction of the period demands from $t-(SI_i^f+T_i^f)$ to t. Again, only those up to $t-S_i$ are relevant. Consequently, we find the coverage random variable to be given by the entire demands of periods $t-(SI_i^f+T_i^f-1)$ to $t-S_i$ and the δ_i^s-fraction of the demands of periods $t-(SI_i^s+T_i^s-1)$ to $t-(SI_i^f+T_i^f)$.

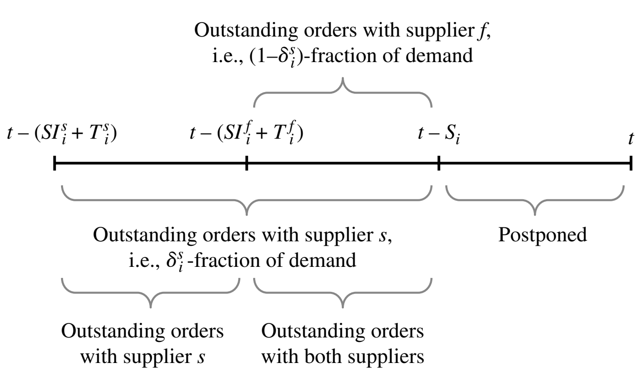
Under stationary conditions $t\to\infty$, the coverage random variable is

$$\tilde{D}_i(SI_i^f, SI_i^s, S_i, \delta_i^s) = D_i(SI_i^f+T_i^f-S_i) + \delta_i^s D_i(SI_i^s+T_i^s-SI_i^f-T_i^f). \tag{9}$$

The net stock, using $\mathbb{E}$ for the expected value and VAR as the variance of a random variable, is

$$IL_i = B_i - \tilde{D}_i(SI_i^f, SI_i^s, S_i, \delta_i^s) \quad \text{with} \tag{10}$$

Figure 5 **Timeline for the Net Stock Calculation**



$$\mathbb{E}[\tilde{D}_i(SI_i^f, SI_i^s, S_i, \delta_i^s)] = ((SI_i^f+T_i^f-S_i) + \delta_i^s(SI_i^s+T_i^s-SI_i^f-T_i^f))\mu_i, \tag{11}$$

$$\mathrm{VAR}[\tilde{D}_i(SI_i^f, SI_i^s, S_i, \delta_i^s)] = ((SI_i^f+T_i^f-S_i) + [\delta^s]^2(SI_i^s+T_i^s-SI_i^f-T_i^f))\sigma_i^2. \tag{12}$$

Proposition 2. *For $\delta_i^s = 1$, the coverage random variable (9) reduces to $D_i(SI_i^s+T_i^s-S_i)$, which corresponds to the well-known one of the single-supply model, $D_i(SI_i+T_i-S_i)$.*

To avoid confusion, we interpret "single supply" as referring to *two* settings. First, a node with a preceding process that requires only a single item and this item is provided by a single supplier. Second, an assembly process where each of the required components is supplied by a single supplier. Both cases result in the same coverage random variable in the standard GS model: $D_i(SI_i+T_i-S_i)$, where $SI_i = \max\{S_i - T_i, \max\{S_j \mid j\colon (j,i)\in\mathcal{A}\}\}$ is the incoming service time of (assembly) node i representing the time to receive all the required inputs from the suppliers to start processing. As a result of Proposition 2, the coverage random variable for a dual-supply node (9) comprises these two settings as special cases.

4.1.2. Safety Stock and Base-Stock Level. For a given set of incoming and outgoing service times, $\tilde{D}_i(SI_i^f, SI_i^s, S_i, \delta_i^s)$ is completely defined. The optimal base-stock level for a predefined nonstockout probability target, α_i-service level, is the smallest one, for which the following (in)equality holds:

$$\Pr\{\tilde{D}_i(SI_i^f, SI_i^s, S_i, \delta_i^s) \le B_i\} \ge \alpha_i. \tag{13}$$

Denote this value as $B_i(SI_i^f, SI_i^s, S_i, \delta_i^s, \alpha_i)$. Then, the safety stock is found as

$$\begin{aligned} R_i(SI_i^f, SI_i^s, S_i, \delta_i^s, \alpha_i) &= B_i(SI_i^f, SI_i^s, S_i, \delta_i^s, \alpha_i) - \mathbb{E}[\tilde{D}_i(SI_i^f, SI_i^s, S_i, \delta_i^s)] \\ &= k_i(\alpha_i)\cdot\sigma_i\cdot\sqrt{(SI_i^f+T_i^f-S_i)+[\delta_i^s]^2(SI_i^s+T_i^s-SI_i^f-T_i^f)} \end{aligned} \tag{14}$$

since we assume normally independent and identically distributed period demand, and thus $\tilde{D}_i(SI_i^f, SI_i^s, S_i, \delta_i^s)$ also follows a normal distribution. The safety factor is denoted by $k_i(\alpha_i)$, which depends on the predefined service-level target α_i and is assumed to be larger than 0 such that the planned safety stock level is nonnegative.

4.1.3. Holding Cost. We follow the usual convention and model the holding cost per unit per period as the foregone interest on an item's (procurement) cost. A node with a single supplier for each item faces a single procurement cost. Thus, the computation is straightforward. A node with two suppliers may face different procurement costs from the different suppliers. Here, the relevant procurement cost for the holding-cost computation is made up of both cost components. Because replenishment orders follow a static allocation, each supplier provides a fixed-supply fraction. The inventory composition also follows this ratio and thus the relevant procurement cost as well. Let $g_{j,i}$ denote the cost added to an item when proceeding from node j to i and c_i denote the cumulative cost of an item at node i with $c_i = 0$ for $i \leq 0$. Then, the holding cost per unit and period is

$$h_i = \nu \cdot c_i = \nu \cdot \sum_{(j,i)\in\mathcal{A}} \delta_{j,i}(c_j + g_{j,i}), \tag{15}$$

where ν denotes the holding-cost/interest rate for the underlying base period, e.g., one week.

4.2. Multinode Model and Analysis with Static Allocation

We first state the entire optimization problem for finding the optimal service times in general acyclic networks and provide the explanation of the objective function and the constraints afterward:

$$\mathbf{P} \quad \text{minimize} \quad C = \sum_{i=1}^{n} h_i \cdot k_i(\alpha_i) \cdot \sigma_i \cdot \sqrt{(SI_i^f + T_i^f - S_i) + [\delta_i^s]^2 (SI_i^s + T_i^s - SI_i^f - T_i^f)} \tag{16}$$

s.t.

$$SI_i^s + T_i^s \geq SI_i^f + T_i^f, \quad i = 1, \ldots, n; \tag{17}$$

$$SI_i^f + T_i^f \geq S_i, \quad i = 1, \ldots, n; \tag{18}$$

$$SI_i^s + T_i^s \geq S_j + T_{j,i}, \quad i = 1, \ldots, n, \forall (j,i) \in \mathcal{A}; \tag{19}$$

$$S_j + T_{j,i} - (SI_i^f + T_i^f) \leq Z \cdot y_{j,i}, \quad i = 1, \ldots, n, \forall (j,i) \in \mathcal{A}; \tag{20}$$

$$\sum_{(j,i)\in\mathcal{A}} y_{j,i} = 1, \quad i = 1, \ldots, n; \tag{21}$$

$$y_{j,i} \in \{0,1\}, \quad i = 1, \ldots, n, \forall (j,i) \in \mathcal{A}; \tag{22}$$

$$T_i^f \geq (1 - y_{j,i}) \cdot T_{j,i}, \quad i = 1, \ldots, n, \forall (j,i) \in \mathcal{A}; \tag{23}$$

$$SI_i^f \geq (1 - y_{j,i}) \cdot S_j, \quad i = 1, \ldots, n, \forall (j,i) \in \mathcal{A}; \tag{24}$$

$$\delta_i^s = \sum_{(j,i)\in\mathcal{A}} y_{j,i} \cdot \delta_{j,i}, \quad i = 1, \ldots, n; \tag{25}$$

$$T_i^s = \sum_{(j,i)\in\mathcal{A}} y_{j,i} \cdot T_{j,i}, \quad i = 1, \ldots, n; \tag{26}$$

$$S_i \geq 0, \quad i = 1, \ldots, n; \tag{27}$$

$$S_i \leq s_i, \quad i \in E; \tag{28}$$

$$S_i = 0, \quad i = 0, -1, \ldots, \tag{29}$$

where s_i is the given maximum service time for demand node $i \in E$ and Z is a sufficiently large number. The objective of **P** is to minimize the total safety stock holding cost in the entire supply network. Pipeline inventory costs are not included because they do not depend on the service times. Constraints (17) and (18) ensure the replenishment lead time and service-time relationships as given in (7). For the explanation of the other constraints, we introduce the following lemma.

Lemma 1. *The objective function* (16) *increases in the slow and fast replenishment lead times,* $SI_i^s + T_i^s$ *and* $SI_i^f + T_i^f$.

Constraints (20) set the slow replenishment lead time of node i, $SI_i^s + T_i^s$, equal to the maximum over the replenishment lead times from all supplying nodes j, $\forall (j,i) \in \mathcal{A}$, because in an optimal solution the equality will hold for (at least) one of the constraints as a result of Lemma 1.

Constraints (20)–(22) ensure two things: (i) The binary variable $y_{j,i}$ is only set equal to 1 if node j represents the slow supplier of node i. (ii) The fast replenishment lead time, $SI_i^f + T_i^f$, is set equal to the second largest replenishment lead time of node i. The reasoning is as follows.

Lemma 2. *Constraints* (20) *ensure that* $y_{j,i} = 1$ *for a node* j *that causes the largest replenishment lead time for node* i.

Constraint (21) ensures that $y_{j,i} = 1$ for not more than one predecessor j of node i. Consequently, constraints (20) also ensure that $SI_i^f + T_i^f$ is set equal to the second largest replenishment lead time of node i because of Lemma 1 and the fact that $SI_i^f + T_i^f$ can only be smaller than one $S_j + T_{j,i}$, namely, the largest one because of Lemma 2. The differentiation between the slow and fast supplier is only relevant for the stock computation at a dual-sourced node, i.e., a node with two predecessors. Here, the predecessor that causes the second largest replenishment lead time directly corresponds to the fast supplier. At single-sourced (assembly) nodes the largest replenishment lead time, which corresponds to the slow replenishment lead time, needs to be covered anyway. The fast replenishment lead time does not play a role because of Proposition 2 since $\delta_i^s = 1$. Therefore, we can identify the fast processing and incoming service time by (23) and (24) because of Lemma 1. The slow supply fraction and processing time are found by (25) and (26).

The remaining constraints ensure that the service times are nonnegative (27), the demand nodes satisfy their service guarantees (28), and the external suppliers deliver the items immediately (29).

Note that the definition of the incoming service times (6) is guaranteed by the constraints of **P** in an optimal solution because (20) together with (23) and (24) implies that $SI_i^f \geq S_i^f$, $i = 1, \ldots, n$ and $SI_i^f \geq S_i - T_i^f$ according to (18). As a result of Lemma 1, in an optimal solution, the equality will hold for (at least) one of the constraints, which ensures $SI_i^f = \max\{S_i - T_i^f,\ S_i^f\}$. The same holds for SI_i^s because of constraints (20) together with (26) and (17) together with (18). Moreover, this implies that SI_i^s and SI_i^f are nonnegative and thus no additional constraints need to be introduced.

With regard to the optimal service times and the computational complexity of **P**, we find the following:

Proposition 3. *For normally distributed demand and a nonstockout probability service-level constraint*, (i) *an extreme point property holds for* **P** *and* (ii) **P** *is NP-hard.*

As a result of Proposition 3, we know that an optimum for **P** is at an extreme point of the feasible region. Exploiting this finding can help in the development of optimization algorithms similar to the dynamic programming algorithms for the single-supply problem for serial, divergent, or convergent supply networks (see Minner 1997).

5. Application

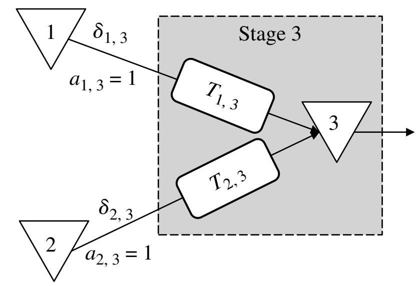
We use the developed model to determine strategic safety stocks in the real-world supply network depicted in Figure 1. We compare the result of our exact approach with the approximate one outlined in the final section of Graves and Willems (2005).

5.1. Approximate Approach (see Graves and Willems 2005)

The standard GS model employs an activity-on-node network representation (see §3.1). Because a single index value can uniquely identify every potential stocking location and its *single* preceding process, it combines the two into a so-called stage. In a multisourcing setting, no equivalent mapping exists. The subtle, but critical, issue is that a potential stocking location supplied by dual sourcing will see two service times and two transportation times, one from each supplier. These individual times can, and almost surely will, differ, so it is not possible to accurately model these times with a single receiving stage. Recognizing this issue, the approximate solution outlined in Graves and Willems (2005) replaces each multisourced stage with one substage for every supplier plus a succeeding pseudo assembly stage with a processing time of 0.

For illustration purposes, Figures 6 and 7 depict the same supply chain where the material at node 3 is dual sourced from nodes 1 and 2 with supply fractions $\delta_{1,3}$ and $\delta_{2,3}$. The transportation processes take $T_{1,3}$ and $T_{2,3}$. Because items are only transported (and not assembled), the production coefficients are $a_{1,3} = a_{2,3} = 1$. Figure 6 is our multisourcing network representation, allowing the exact inventory calculations

Figure 6 **Stage Assignment—Exact Model**

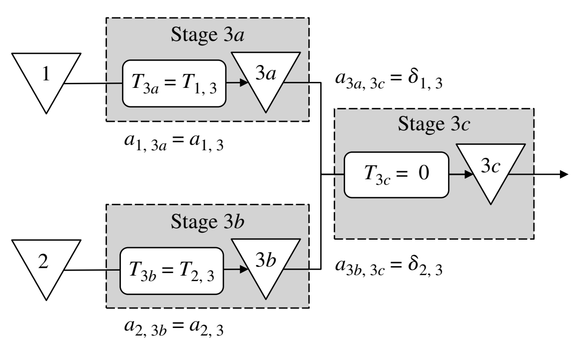


presented in §4. Figure 7 is the approximate remodeling, which is obtained as follows:

1. The dual-sourced stage 3 is split into three substages, i.e., stages $3a$, $3b$, and $3c$.
2. Stages $3a$ and $3b$ are associated with the transportation processes from the supply nodes and thus, mapping the standard GS single index-value notation to our notation, processing times $T_{3a} = T_{1,3}$ and $T_{3b} = T_{2,3}$. The production coefficients of these stages and their supplying ones are $a_{1,3a} = a_{2,3b} = 1$.
3. The split of the original demand at stage 3 according to the supply fractions is achieved by introducing a pseudo assembly stage $3c$ with processing time $T_{3c} = 0$ directly after the transportation processes. The supply fractions, $\delta_{1,3}$ and $\delta_{2,3}$, which do not exist in the standard GS approach, are turned into the production coefficients $a_{3a,3c} = \delta_{1,3}$ and $a_{3b,3c} = \delta_{2,3}$, respectively.
4. In terms of cost, the inventory holding costs per unit and period at stages $3a$, $3b$, and $3c$ are set equal to the holding cost at stage 3 in the original network.

In this remodeling process the approximation takes place in step 3. Even though the split of the original demand across the two suppliers can be done correctly by setting the production coefficients equal to the supply fractions, the resulting inventory calculations at the three substages cannot accurately capture the inventory pooling effect. Because the original dual-sourced node 3 is represented by three substages constituting an assembly operation, stage $3c$ assumes that stages $3a$ and $3b$ deliver *different components* for its assembly process. To start the assembly process, both input

Figure 7 **Stage Assignment—Workaround in the Standard GS Approach**



components from stages $3a$ and $3b$ need to be available *at the same point in time* and in the required ratio. That is why stage $3c$ considers only a single incoming service time, which is the maximum of both outgoing service times of the supplying stages, S_{3a} and S_{3b}, when determining its inventory requirement, instead of dealing with these service times individually. This results in the following approximate cost function at dual-sourced node 3 (assuming normally distributed demand, $S_1 + T_{1,3} > S_2 + T_{2,3}$ without loss of generality, and the constraints on the service times specified in **P**):

$$\tilde{C}(S_1, S_2, S_{3a}, S_{3b}, S_{3c}) = \begin{cases} h_3 k_3 \sigma_3 \Big(\delta_{1,3}\sqrt{S_1+T_{1,3}-S_{3a}} + \delta_{2,3}\sqrt{S_2+T_{2,3}-S_{3b}} \\ \qquad + \sqrt{\max\{S_{3a}, S_{3b}\} - S_{3c}} \Big) & 0 \le S_{3c} < S_2 + T_{2,3}, \\ h_3 k_3 \sigma_3 \Big(\delta_{1,3}\sqrt{S_1+T_{1,3}-S_{3a}} + \sqrt{S_{3a}-S_{3c}} \Big) & S_2+T_{2,3} \le S_{3c} < S_1+T_{1,3}. \end{cases} \tag{30}$$

Note that if $S_{3c} \ge S_2 + T_{2,3}$, no stock needs to be kept at stage $3b$ because it can act in a just-in-time fashion. Hence, this term vanishes in the cost function. With regard to the optimal stock positioning at a dual-sourced stage/node that is remodeled according to this approximation, we find the following:

Lemma 3. *At a dual-sourced node, which is approximately remodeled as an assembly system with three substages, it is optimal under normally distributed demand to set the service times of the two upstream stages equal to the service time of the downstream stage. Hence, stock is never held at the downstream stage.*

In reality, stages $3a$ and $3b$ do not provide different components but the *very same item*. Therefore, the items do not need to arrive at the same point in time but may well arrive at *different points in time*. It only needs to be ensured in the stock computation at stage $3c$ that one of the suppliers has the item available at the requested point in time, not both. This is reflected by the cost function of the exact approach, which is (according to (14) and Proposition 1)

$$C(S_1, S_2, S_3) = \begin{cases} h_3 k_3 \sigma_3 \\ \quad \cdot \sqrt{(S_2+T_{2,3}-S_3) + \delta_{1,3}^2 (S_1+T_{1,3}-(S_2+T_{2,3}))} & 0 \le S_3 < S_2+T_{2,3}, \\ h_3 k_3 \sigma_3 \sqrt{\delta_{1,3}^2 (S_1+T_{1,3}-S_3)} & S_2+T_{2,3} \le S_3 \le S_1+T_{1,3}. \end{cases} \tag{31}$$

The remodeling of the supply network in the approximate approach only affects dual-sourced nodes. At the other nodes, both the exact and the approximate approaches determine the required safety stock level according to the same logic. Based on (30) and (31) and Lemma 3, we find the following:

Proposition 4. *For normally distributed demand and identical replenishment lead times at a dual-sourced node in both the exact and the approximate approach, the exact approach only outperforms the approximate one if the dual-sourced node's outgoing service time is smaller than the shorter replenishment lead time.*

The inferior performance of the approximation is caused by an overestimation of the required stock level at a dual-sourced node if its outgoing service time is smaller than the shorter replenishment lead time. As a result of the remodeling into an assembly stage, the approximate approach cannot accurately capture the inventory pooling potential. It holds two separate safety stocks for the items provided by the two suppliers because it views them as different components. This is reflected by the two separate square root expressions in the optimal solution to the first case of cost function (30), which becomes after substituting $S_{3a} = S_{3b} = S_{3c} = S_3$ because of Lemma 3:

$$\tilde{C}(S_1, S_2, S_3) = h_3 k_3 \sigma_3 \Big(\sqrt{\delta_{1,3}^2 (S_1 + T_{1,3} - S_3)} + \sqrt{(1-\delta_{1,3})^2 (S_2 + T_{2,3} - S_3)} \Big). \tag{32}$$

The first case of the exact cost function (31) can be rewritten as

$$C(S_1, S_2, S_3) = h_3 k_3 \sigma_3 \sqrt{\delta_{1,3}^2 (S_1+T_{1,3}-S_3) + (1-\delta_{1,3}^2)(S_2+T_{2,3}-S_3)}. \tag{33}$$

This case is easily shown to be smaller than (32) (see proof of Proposition 4). This overestimation of the required stock level in the approximate approach may not only lead to different safety stock levels prescribed by the two approaches at identical stock-holding nodes in the network but also to a completely different safety stock positioning across the network (as we will see in §5.3).

5.2. Optimization Algorithms

After remodeling the original supply network, the standard GS dynamic programming (DP) algorithm (see, e.g., Graves and Willems 2000) can be applied to obtain the optimal solution to the approximation. For the exact approach, the real-world supply network has an assembly structure, allowing us to adjust the standard DP algorithm such that it can also handle dual-sourced nodes. In essence, we formulate two cost functions, one for a single-sourced node and one for a dual-sourced one. We obtain the optimal solution by a forward recursion; see the online appendix for algorithmic details.

5.3. Performance Comparison: Approximate vs. Exact Approach

To illustrate and analyze the differences between the exact and the approximate approach, we start off by applying both models to the real-world example in Figure 1.

5.3.1. Data. Table 1 shows the processing times, per unit added costs, and supply fractions for the company's current supply chain. Note that the data is disguised to protect confidential company data but the underlying relations remain true. For this reason, we also simply choose $\nu = 1$. Although China has significant labor savings compared with Texas, the material costs are more closely matched. Transportation costs are more expensive from China because it must ship by sea and truck. All production coefficients $a_{i,j}$ are equal to 1. Disguised demand on a daily basis is normally distributed with a mean of 100 and a sigma of 40. At all locations, a safety factor of 2.33 is assumed, which corresponds to an α-service level of 99%. The maximum allowed outgoing service time for the external demand node 10 is assumed to be 0, i.e., $s_{10} = 0$.

5.3.2. Results. Figure 8 presents the optimal service times and stocking locations of the exact approach resulting in a total safety stock cost of \$3,715.98. Safety stock is held at dark-shaded nodes. By definition, node 10 satisfies external customer demand and holds safety stock. In addition, the two most upstream nodes and the dual-sourced node 6 hold safety stock in order to synchronize the outgoing service times quoted to assembly node 9.

Table 1 Supply Network Parameters

Arc (i, j)	Processing time (in days), $T_{i,j}$	Added cost (in \$), $g_{i,j}$	Supply fraction, $\delta_{i,j}$
(0, 1)	20	0.3	1
(0, 2)	20	0.25	1
(1, 3)	5	2	1
(2, 3)	5	2	1
(0, 4)	20	3.5	1
(0, 5)	3	1.05	0.4
(−1, 5)	10	0.95	0.6
(3, 6)	0	0	0.75
(0, 6)	20	3	0.25
(0, 7)	7	0.8	1
(4, 8)	5	2.6	1
(5, 8)	5	2.6	1
(6, 9)	30	0.9	1
(7, 9)	30	0.9	1
(8, 10)	3	0.1	0.3
(9, 10)	30	0.25	0.7

The remodeled supply network representation from Figure 8 is presented in Figure 9. The stages surrounded by the light-grey background illustrate originally dual-sourced nodes that are replaced by substages to make the supply network fit into the standard GS model.

In the optimal solution of the approximate approach, safety stock is held at the same stages as in the exact approach, which is illustrated by the dark-grey shading in Figure 9. All optimal service times are identical to the ones of the exact approach. However, the safety stock levels and thus the total cost are different because the approximate approach cannot coordinate the replenishment lead times at dual-sourced nodes correctly and thus cannot properly capture inventory pooling. At node 10, a larger safety stock (1,155.95 units) is held versus the exact approach (1,113.68 units), which causes the total safety stock cost to increase by 6.13% to \$3,943.63.

5.3.3. Further Analysis. To generalize these results, we use our example as the basis for a numerical exercise where we allow each dual-sourced node to assume a supply fraction of 0.25, 0.50 or 0.75. Because there are three dual-sourced nodes, this results in 27 instances. All other input parameters are the same as the original data, although the change in the supply fractions obviously changes both the demands and costs across each problem instance.

Tables 2–4 present the results comparing the exact (E) and the approximate (A) approach. The first three columns indicate the supply fractions at the dual-sourced nodes 5, 6, and 10, which define the specific instance. In 10 of the 27 instances, the optimal safety stock policy (i.e., optimal service times) differs between the two approaches. Safety stock is held at different nodes in the supply network. These instances are listed first in Table 2, where for each node/stage (columns 1–9) the optimal outgoing service time of the two approaches (columns E and A) is summarized. Because the approximation remodels each of the dual-sourced nodes 5 and 6 by three substages, of which only the outgoing service times of the two upstream stages are relevant because of Lemma 3, we report two outgoing service times under column A at those nodes. Moreover, node 10 is not included in the table because it quotes an optimal outgoing service time of 0 in all instances due to $s_{10} = 0$. The last column indicates whether the service-time pattern is identical to or different from the exact and the approximate approach.

Table 3 provides the safety stock levels at the different nodes resulting from the prescribed service times for those instances where the service-time patterns differ between the two approaches. For example, in the first two instances the exact approach (E) holds safety stock at dual-sourced nodes 5 and 10 plus

Figure 8 Optimal Safety Stock Placement in the Figure 1 Multisourced Supply Chain

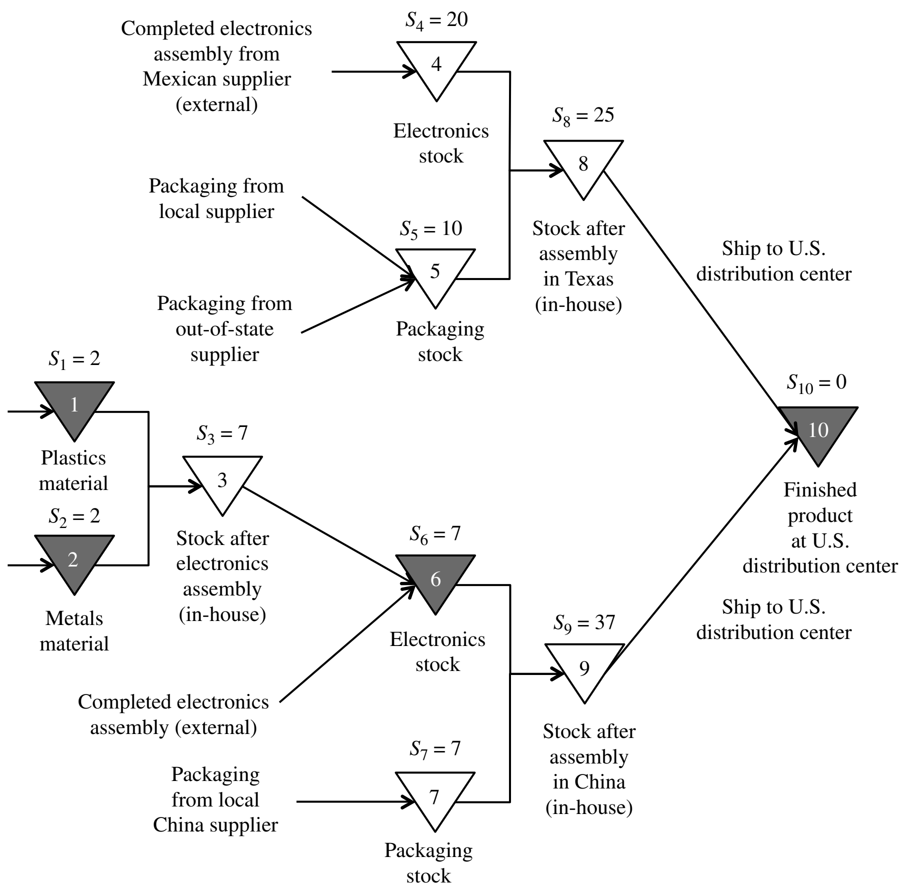


Figure 9 Optimal Solution in the Approximate Remodeling of the Figure 1 Multisourced Supply Chain

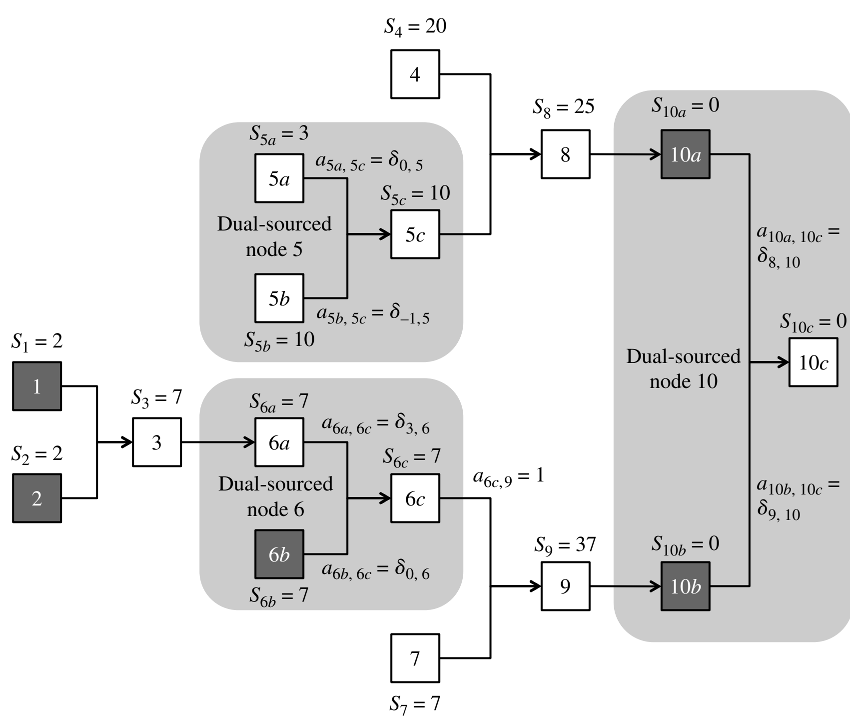


Table 2 Service-Time Patterns of the Exact (E) and Approximate (A) Approaches

Supply fraction			Service time at node/stage																			Service-time patterns of E and A
			1		2		3		4		5			6			7	8		9		
												A			A							
0 to 5	3 to 6	8 to 10	E	A	E	A	E	A	E	A	E	5a	5b	E	6a	6b	E/A	E	A	E	A	
0.75	0.75	0.75	15	2	15	2	20	7	0	20	0	3	10	20	7	7	7	5	25	50	37	Different
0.50	0.75	0.75																				
0.75	0.25	0.75	15	15	15	15	20	20	0	20	0	3	10	20	20	20	7	5	25	50	50	
0.75	0.50	0.75																				
0.50	0.25	0.75																				
0.50	0.50	0.75																				
0.25	0.75	0.75	20	2	20	2	25	7	20	20	10	3	10	25	7	7	7	25	25	55	37	
0.25	0.75	0.50	15	2	15	2	20	7	20	20	10	3	10	20	7	7	7	25	25	50	37	
0.50	0.75	0.50																				
0.75	0.75	0.50																				
0.75	0.25	0.50	15		15		20		20		10	3	10	20	20	20	7	25		50		Identical
0.50	0.25	0.50																				
0.25	0.25	0.50																				
0.75	0.50	0.50																				
0.50	0.50	0.50																				
0.25	0.50	0.50																				
0.25	0.25	0.75																				
0.25	0.50	0.75																				
0.75	0.25	0.25																				
0.50	0.25	0.25																				
0.25	0.25	0.25																				
0.75	0.50	0.25																				
0.50	0.50	0.25																				
0.25	0.50	0.25																				
0.75	0.75	0.25	2		2		7		20		10	3	10	7	7	7	7	25		37		
0.50	0.75	0.25																				
0.25	0.75	0.25																				

further stocks at single-sourced nodes 1, 2, and 4. The approximate approach (A) carries safety stock at dual-sourced nodes 6 and 10 in addition to stock at single-sourced nodes 1 and 2. Even though the total stock levels across all nodes are 14%–16% lower in the approximation, the total cost is about 10% higher. Across all instances in Table 3, the average total cost of the approximation is 10.63% above the optimal exact cost. The maximum difference is 12.34%.

These instances show that we need an integrated approach like the newly developed exact approach in this paper. The reason is that in the process of determining the optimal stock-holding locations, each of the two approaches trades off the following two effects: (i) the *risk pooling effect* over time, which advocates holding more stock downstream, because the required additional safety stock level for buffering against an additional time unit is decreasing; and (ii) the *value-added effect* through further processing, which makes inventory holding costs higher at downstream nodes. Because of the remodeling into an assembly stage, the approximate approach basically considers a different risk-pooling effect than the exact approach. Therefore, it solves a different trade-off within the optimization. At node 10 (but also at the other dual-sourced nodes), this becomes quite obvious. Because of the remodeling, the approximation treats the upper supply network part with stage 10a and all its suppliers completely independent of the lower part with stage 10b and its suppliers. This results in different cost trade-offs and thus may also result in a different "optimal" safety stock policy under both approaches as illustrated in Table 3.

Table 4 summarizes the safety stock levels of those instances where the exact and the approximate approaches find identical optimal service times (see Table 2). Two different service-time patterns are obtained; one where node 10 is the only dual-sourced node holding stock (instances 1–14 in Table 4) and one where dual-sourced nodes 6 and 10 hold stock (besides further single-sourced nodes). In both patterns the cost difference between the two approaches arises solely from the different stock levels prescribed at node 10. Even though the second service-time pattern holds stock at dual-sourced node 6 as well, the exact approach cannot realize any benefits over the approximate one because the optimal service time is *not* smaller than the shorter replenishment lead time

Table 3 **Different Service-Time Patterns—Safety Stock Quantity and Cost Comparison Between the Exact (E) and Approximate (A) Approaches**

Supply fraction			Safety stock at node/stage															Total stock diff. (%)	Total cost diff. (%)
			1		2		4		5			6			10				
										A			A			A			
0 to 5	3 to 6	8 to 10	E	A	E	A	E	A	E	5a	5b	E	6a	6b	E	10a	10b		
0.75	0.75	0.75	39.08	74.14	39.08	74.14	312.60	0	129.60	0	0	0	0	21.00	329.51	369.88	190.72	−14.12	10.75
0.50	0.75	0.75	39.08	74.14	39.08	74.14	312.60	0	152.34	0	0	0	0	21.00	329.51	369.88	190.72	−16.36	10.01
0.75	0.25	0.75	13.03	13.03	13.03	13.03	312.60	0	129.60	0	0	0	0	0	329.51	369.09	208.40	−24.35	12.34
0.75	0.50	0.75	26.05	26.05	26.05	26.05	312.60	0	129.60	0	0	0	0	0	329.51	369.88	208.40	−23.48	12.15
0.50	0.25	0.75	13.03	13.03	13.03	13.03	312.60	0	152.34	0	0	0	0	0	329.51	369.88	208.40	−26.35	11.59
0.50	0.50	0.75	26.05	26.05	26.05	26.05	312.60	0	152.34	0	0	0	0	0	329.51	369.88	208.40	−25.54	11.39
0.25	0.75	0.75	0	74.14	0	74.14	0	0	0	0	0	0	0	21.00	523.60	369.88	190.72	39.40	9.89
0.25	0.75	0.50	78.15	148.28	78.15	148.28	0	0	0	0	0	0	0	42.00	596.77	246.58	381.44	28.35	9.40
0.50	0.75	0.50																	9.39
0.75	0.75	0.50																	9.38

(see Proposition 4). Hence, both approaches prescribe an identical safety stock level of 63.01 units.

At dual-sourced node 10 the situation is different. The service time is 0 and thus smaller than the shorter replenishment lead time. The approximate approach prescribes safety stock levels, which are 3%–10% higher than the exact approach. This overestimation causes an average and maximum total cost increase of 8.27% and 11.12%. Consequently, we observe that even though the approximation might find the optimal service-time pattern and thus stocking locations, the correct sizing of the safety stock levels at dual-sourced nodes can only be done according to (14) of the exact approach.

6. Conclusion

In this paper, we have integrated static dual supply into the guaranteed-service framework. The developed model characterizes inventory requirements exactly in dual-supply general acyclic supply networks where the suppliers are allocated static fractions of demand. Compared with the only previously published approximate solution, our exact approach may result in considerable cost savings. In our 27-instance numerical study, total safety stock cost savings amount to 9.1%, on average. These savings are realized because the exact model captures inventory pooling in a way that the approximation is unable to do. Consequently, the approximate

Table 4 **Identical Service-Time Patterns—Safety Stock Quantity and Cost Comparison Between the Exact (E) and Approximate (A) Approaches**

Supply fraction			Safety stock at node/stage								Total stock diff. (%)	Total cost diff. (%)
			1	2	6			10				
						A			A			
0 to 5	3 to 6	8 to 10	E/A	E/A	E	6a	6b	E	10a	10b		
0.75	0.25	0.50	26.05	26.05	0	0	0	596.77	246.58	416.80	10.27	11.12
0.50	0.25	0.50										
0.25	0.25	0.50										
0.75	0.50	0.50	52.10	52.10	0	0	0	596.77	246.58	416.80	9.50	11.07
0.50	0.50	0.50										
0.25	0.50	0.50										
0.25	0.25	0.75	13.03	13.03	0	0	0	521.00	369.88	208.40	10.47	10.97
0.25	0.50	0.75	26.05	26.05	0	0	0	521.00	369.88	208.40	9.99	10.95
0.75	0.25	0.25	39.08	39.08	0	0	0	705.19	123.29	625.20	5.53	6.11
0.50	0.25	0.25										
0.25	0.25	0.25										
0.75	0.50	0.25	78.15	78.15	0	0	0	705.19	123.29	625.20	5.03	6.07
0.50	0.50	0.25										
0.25	0.50	0.25										
0.75	0.75	0.25	222.42	222.42	63.01	0	63.01	658.61	123.29	572.16	3.16	5.16
0.50	0.75	0.25										
0.25	0.75	0.25										

approach may place safety stock at different locations than the exact one. Although the total safety stock quantity in the entire supply network does not necessarily have to be higher (it may even be lower in some cases), the total costs resulting from such a different safety stock allocation in the network are higher. Even if the approximate approach finds the optimal safety stock locations, it may still prescribe higher safety stock levels at these locations than is actually required, thus causing higher costs. Also in those cases the exact model is needed for rightsizing the stock level.

This research raises several relevant questions for further consideration. We conclude with three issues that we think are most worthy of additional work. First, efficient solution algorithms enabling the optimization of large general acyclic networks with dual supply should be developed. For the solution of the practical example in this paper, which exhibits a pure assembly structure, we have presented a dynamic programming algorithm. Other network structures require a modification of this algorithm or a completely different solution approach. Starting points could be the branch-and-bound algorithm of Humair and Willems (2011) or the mixed-integer programming formulation of Magnanti et al. (2006).

Second, relaxing the assumption of static, exogenously determined supply fractions in multiechelon networks deserves investigation. The single-echelon analysis by Boute and van Mieghem (2013) shows substantial cost differences with an order-splitting policy with dynamically adjusted supply fractions in certain parameter settings. Compared with our strategic model, this analysis and most other single-echelon models have an operational focus and thus only consider the trade-off between procurement cost and inventory holding cost. At a strategic level, as mentioned in our paper, the determination of the optimal supply fractions requires the consideration of additional aspects like risk mitigation and securing upside volume. Therefore, the current model would have to be extended in these respects when turning the supply fractions into decision variables. Because many single-echelon models do not address these aspects either, one could start with such an analysis first. Besides a relaxation of the static supply fraction assumption in the order-splitting policy, the integration of other replenishment policies in a multiechelon setting, such as the constant-order or dual-index policy, represents a valuable extension as well.

Third, the development of a multiechelon multisupply model under the stochastic-service assumption is worth exploring. In our model we assume guaranteed service, i.e., a stage provides 100% service for its quoted service time. However, in large parts of the inventory literature it is assumed that the service time between nodes can vary depending on the material availability at the upstream node. In the single-echelon setting, several contributions exist that consider dual sourcing under stochastic replenishment lead times, but hardly any address the multiechelon context.

Supplemental Material

Supplemental material to this paper is available at http://dx.doi.org/10.1287/msom.2013.0472.

Acknowledgments

The authors thank the editors and three referees for their very helpful and constructive feedback on earlier versions of this paper.

References

Aggarwal PK, Moinzadeh K (1994) Order expedition in multi-echelon production/distribution systems. *IIE Trans.* 26(2): 86–96.

Allon G, van Mieghem JA (2010) Global dual sourcing: Tailored base surge allocation to near and offshore production. *Management Sci.* 56(1):110–124.

Arts J, van Vuuren M, Kiesmüller GP (2011) Efficient optimization of the dual-index policy using Markov Chains. *IIE Trans.* 43(8):604–620.

Berling P, Martínez-de-Albéniz V (2012) Optimal expediting decisions in a continuous-stage serial supply chain. Research report, IESE Business School, University of Navarra, Barcelona, Spain.

Billington C, Callioni G, Crane B, Ruark JD, Rapp JU, White T, Willems SP (2004) Accelerating the profitability of Hewlett-Packard's supply chains. *Interfaces* 34(1):59–72.

Boute RN, van Mieghem JA (2013) Global dual sourcing and order smoothing: The impact of capacity and leadtime. Working paper, Kellogg School of Management, Northwestern University, Evanston, IL.

Chiang C (2007) Optimal control policy for a standing order inventory system. *Eur. J. Oper. Res.* 182(2):695–703.

Clark AJ, Scarf H (1960) Optimal policies for a multi-echelon inventory problem. *Management Sci.* 6(4):465–490.

Daniel KH (1962) A delivery-lag inventory model with emergency order. Scarf H, Gilford D, Shelly M, eds. *Multi-Stage Inventory Models and Techniques*, Chap. 2 (Stanford University Press, Stanford, CA).

Farasyn I, Humair S, Kahn JI, Neale JJ, Rosen O, Ruark J, Tarlton W, Van de Velde W, Wegryn G, Willems SP (2011) Inventory optimization at Procter & Gamble: Achieving real benefits through user adoption of inventory tools. *Interfaces* 41(1):66–78.

Fukuda Y (1964) Optimal policy for the inventory problem with negotiable leadtime. *Management Sci.* 10(4):609–708.

Ganeshan R (1999) Managing supply chain inventories: A multiple retailer, one warehouse, multiple supplier model. *Internat. J. Production Econom.* 59(1–3):314–354.

Graves SC, Willems SP (2000) Optimizing strategic safety stock placement in supply chains. *Manufacturing Service Oper. Management* 2(1):68–83.

Graves SC, Willems SP (2003) Supply chain design: Safety stock placement and supply chain configuration. de Kok AG, Graves SC, eds. *Supply Chain Management: Design, Coordination, and Operation*, Chap. 3, Handbooks in Operations Research and Management Science (Elsevier, Amsterdam).

Graves SC, Willems SP (2005) Optimizing the supply chain configuration for new products. *Management Sci.* 51(8):1165–1180.

Graves SC, Willems SP (2008) Strategic inventory placement in supply chains: Non-stationary demand. *Manufacturing Service Oper. Management* 10(2):278–287.

Humair S, Willems SP (2011) Technical note: Optimizing strategic safety stock placement in general acyclic networks. *Operations Res.* 59(3):781–787.

Janssen FBSLP, de Kok AG (1999) A two-supplier inventory model. *Internat. J. Production Econom.* 59(1–3):395–403.

Klosterhalfen S (2010) Multiple sourcing in single- and multi-echelon inventory systems. Ph.D. thesis, Business School, University of Mannheim, Mannheim, Germany. https://ub-madoc.bib.uni-mannheim.de/2959/.

Lawson DG, Porteus EL (2000) Multistage inventory management with expediting. *Oper. Res.* 48(6):878–893.

Magnanti TL, Shen Z-JM, Shu J, Simchi-Levi D, Teo C-P (2006) Inventory placement in acyclic supply chain networks. *Oper. Res. Lett.* 34(2):228–238.

Manary MP, Willems SP (2008) Setting safety stock targets at Intel in the presence of forecast bias. *Interfaces* 38(2):112–122.

Minner S (1997) Dynamic programming algorithms for multi-stage safety stock optimization. *OR Spektrum* 19(4):261–271.

Minner S (2003) Multiple-supplier inventory models in supply chain management: A review. *Internat. J. Production Econom.* 81–82(1):265–279.

Muckstadt JA, Thomas LJ (1980) Are multi-echelon inventory methods worth implementing in systems with low-demand-rate items? *Management Sci.* 26(5):483–494.

Rosenshine M, Obee D (1976) Analysis of a standing order inventory system with emergency orders. *Oper. Res.* 24(6):1143–1155.

Schoenmeyr T, Graves SC (2009) Strategic safety stocks in supply chains with evolving forecasts. *Manufacturing Service Oper. Management* 11(4):657–673.

Sheopuri A, Janakiraman G, Seshadri S (2010) New policies for the stochastic inventory control problem with two supply sources. *Oper. Res.* 58(3):734–745.

Simpson KF Jr (1958) In-process inventories. *Oper. Res.* 6(6):863–873.

Thomas DJ, Tyworth JE (2006) Pooling lead-time risk by order splitting: A critical review. *Transportation Res. Part E* 42(4):245–257.

Tian F, Willems SP, Kempf KG (2011) An iterative approach to item-level tactical production and inventory planning. *Internat. J. Production Econom.* 133(1):439–450.

van Houtum GJ (2006) Multi-echelon production/inventory systems: Optimal policies, heuristics, and algorithms. Gray P, Johnson MP, Norman B, Secomandi N, eds. *Models, Methods, and Applications for Innovative Decision Making*, Chap. 7, Tutorials in Operations Research 2006 (INFORMS, Hanover, MD).

Veeraraghavan S, Scheller-Wolf A (2008) Now or later: Dual index policies for capacitated dual sourcing systems. *Oper. Res.* 56(4):850–864.

Whittemore AS, Saunders SC (1977) Optimal inventory under stochastic demand with two supply options. *SIAM J. Appl. Math.* 32(2):293–305.

Willems SP (2008) Data set—Real-world multiechelon supply chains used for inventory optimization. *Manufacturing Service Oper. Management* 10(1):19–23.