M&SOM MANUFACTURING & SERVICE OPERATIONS MANAGEMENT
Vol. 7, No. 4, Fall 2005, pp. 295–318
ISSN 1523-4614 | EISSN 1526-5498 | 05 | 0704 | 0295

informs

DOI 10.1287/msom.1050.0087

Safety Stock Positioning in Supply Chains with Stochastic Lead Times

David Simchi-Levi
Engineering Systems Division and Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, dslevi@mit.edu

Yao Zhao
Department of Management Science and Information Systems, Rutgers University, Newark, New Jersey 08903, yaozhao@andromeda.rutgers.edu

We study the safety stock positioning problem in single-product multistage supply chains with tree network structures, where each stage controls its inventory using an installation continuous-time base-stock policy. External demands follow independent Poisson processes, and unsatisfied demands at each stage are fully backordered. The processing (e.g., production) cycle times and transportation lead times are assumed to be stochastic, sequential, and exogenously determined. We derive recursive equations for the backorder delays (because of stockout) at all stages in the supply chain. Based on the recursive equations, we characterize the dependencies of the backorder delays across different stages in the network, and develop insights into the impact of safety stock positioning in various supply chain topologies. We present approximations and algorithms to coordinate the base-stock levels in these supply chains, so as to minimize systemwide inventory cost subject to meeting certain service-level requirements of the external customers.

Key words: stock positioning; stochastic sequential lead time; recursive equation; backorder delay; tree network structure
History: Received: December 29, 2003; accepted: July 15, 2005. This paper was with the authors 8 months for 3 revisions.

1. Introduction

Many real-world supply chains have complex network structures, which consist of multiple layers of production and distribution facilities. To cope with uncertainties in demand and supply, these supply chains often have many millions of dollars of capital tied up in inventories (Feigin 1999). One important question, of course, is how to best manage inventories in complex multistage supply chains, so as to meet customer expectations with minimum systemwide inventory holding cost.

One way to answer this question is to characterize the optimal inventory policies, which has been discussed extensively in the inventory control literature. We refer the reader to Zipkin (2000), Federgruen (1993), and Porteus (2002) for excellent reviews. Unfortunately, the optimal policy is not known for supply chains with general (e.g., tree) network structures.

In practice, many companies employ simple heuristic policies, such as the installation base-stock policies, to control inventory at each facility (e.g., Lee and Billington 1993, Graves and Willems 2000, and Lin et al. 2000). In an installation policy, each facility only needs the inputs from immediate upstream and downstream facilities, and makes ordering decisions based on its local order and inventory status. An important and challenging question for these companies is how to optimally coordinate the installation policies at all facilities, so as to minimize systemwide inventory holding cost while meeting the end customers' service requirements. In other words, given that each facility is managed autonomously by an installation policy, how can a central planner determine the policy parameters for all facilities in the best possible way?

In this paper, we attempt to address this question for a class of supply chains with tree structure,

where each facility (or equivalent, each stage) uses a continuous-time base-stock policy to control its inventory, external demands follow independent Poisson processes, and the processing (e.g., production) cycle times and transportation lead times are stochastic. These assumptions are generally valid in supply chains that carry expensive items and face low volume but highly uncertain demand, e.g., service parts supply chains; see Sherbrooke (1968), Muckstadt (1973), Graves (1985), and Caglar et al. (2004) for real-world examples. An excellent literature review is provided by Axsäter (1993). As demonstrated by Lin et al. (2000), studies of these models also provide insights and benchmarks, which may help practitioners significantly improve inventory management in other classes of the real-world supply chains.

It is well known that a base-stock policy is not optimal, in general, for supply chains with a tree structure (Zipkin 2000). However, it is widely used in practice partly because it is simple, and therefore easily implementable, and partly because this policy has been proven optimal or close to optimal in some special but important cases. For instance, in serial supply chains with zero setup costs and without capacity constraints, because the installation base-stock policy is equivalent to an echelon base-stock policy under certain initial conditions (Axsäter and Rosling 1993), it is indeed optimal in these cases (Clark and Scarf 1960). In serial systems with zero setup costs but with finite production capacities, the base-stock policy subject to capacity constraints, i.e., the modified base-stock policy, may not be optimal, but the best modified base-stock policy is still close to optimal (Speck and van der Wal 1991, van Houtum et al. 1996). We refer to Parker and Kapuscinski (2004) and Janakiraman and Muckstadt (2004) for structural results on the optimal policies in these systems.

Many authors have studied the installation policies for supply chains with various network topologies, e.g., multistage serial systems (Simpson 1958, Hanssmann 1959, Lee and Zipkin 1992), distribution systems (Sherbrooke 1968, 1986; Axsäter 1993; Graves 1985; Lee and Moinzadeh 1987a, b; Svoronos and Zipkin 1988, 1991; Schwartz et al. 1985). The majority of this research follows the "building-block" approach (Graves 1988); i.e., a building block is typically a processor plus a stock-keeping facility. The processor can be anything from a workstation, a group of workstations, to an entire assembly plant. Assuming each building block operates independently using a simple installation policy, one can first characterize various building blocks, e.g., serial, assembly, or distribution systems, and then identify the links among these building blocks. For excellent literature reviews in this area, we refer the reader to Graves and Willems (2003b), Axsäter (1993), Diks et al. (1996), and Zipkin (2000).

Previous research most closely related to ours includes Lee and Billington (1993), Ettl et al. (2000), and Graves and Willems (2000). Two distinct models are studied: the stochastic-service model and the guaranteed-service model. We refer the reader to Graves and Willems (2003b) for a thorough comparison between the two models. Briefly, in the stochastic-service model, each stage in the supply chain is subject to stochastic delays of orders received from its upstream stages. These delays are due to of stockouts at the upstream facilities. In the guaranteed-service model, it is assumed that if stockouts occur, each stage in the supply chain has external resources other than the on-hand inventory to serve its downstream facilities, so that the service time is always guaranteed.

Lee and Billington (1993) developed a multi-stage inventory model based on the stochastic-service model approach for the Hewlett-Packard DeskJet printer supply chain with the objective of providing tools for the managers to evaluate various stock positioning strategies. Assuming that each stage controls its inventory by an installation periodic-review base-stock policy, the authors developed approximations for the replenishment lead times at all stages of the supply chain. The authors also called for better models to more accurately characterize the possibly correlated input delays in the assembly systems.

The stochastic-service model approach is utilized by Ettl et al. (2000) to analyze supply chains in which each stage controls its inventory with a continuous-time base-stock policy. Ettl et al. (2000) differentiated the *nominal* lead time and *actual* lead time at each stage. The actual lead time will exceed the nominal lead time if the suppliers are out of stock. The authors analyzed the assembly systems by assuming that at most one supplier can be out of stock at any

time. They derived approximations and bounds on the expected backorder delays (because of stockouts) to downstream customers by modeling the replenishment process at each stage as $M^x/G/\infty$ queue. In addition to performance evaluation, Ettl et al. (2000) optimized the total inventory investment, i.e., the sum of expected work in process and finished goods inventory, in the supply chain subject to meeting certain service requirements of the external customers. Using the safety factors (service levels) as decision variables, the authors developed analytic expressions for the gradients, and therefore the constrained nonlinear optimization problem can be solved by the conjugate gradient method.

Graves and Willems (2000) applied the guaranteed-service model approach to tree-structure supply chains. Instead of base-stock levels or service levels, planned lead times at all stages are used as the decision variables. It is assumed that the planned lead time at each stage of the supply chain is 100% guaranteed to the downstream stages, thus the lead time between every two stages is deterministic. To optimize the safety stock placement, Graves and Willems (2000) (see also Graves and Willems 2003a) developed a fast algorithm based on dynamic programming.

This paper follows *the stochastic-service model approach* and assumes that each stage controls its inventory by a continuous-time base-stock policy as in Ettl et al. (2000). However, we make a different assumption on the lead times: while Ettl et al. (2000) modeled the replenishment lead times at all stages by i.i.d. random variables, we make the "transit times" assumption (Svoronos and Zipkin 1991) as follows:

ASSUMPTION 1.1. *The transportation lead times, processing cycle times, and replenishment lead times of external suppliers are stochastic, sequential, and exogenously determined.*

In particular, the "transit time" satisfies the following assumptions: (1) the "transit time" is independent of the system state, e.g., demand and order placement, (2) the "transit time" is independent across stages. Under this assumption, order crossing is not allowed and Palm's theorem (1938) cannot be applied. Svoronos and Zipkin (1991) pointed out that the "transit time" assumption may be more realistic than the i.i.d. lead-time assumption in some real-world applications; see Zipkin (2000) for a thorough discussion of these assumptions.

Under the "transit time" assumption, Svoronos and Zipkin (1991) and Lee and Zipkin (1992, 1995) studied serial and distribution systems; see also Zipkin (2000). However, assembly systems pose substantial challenges because they "are notoriously difficult to analyze" (Lee and Zipkin 1995).

In this paper, we extend the theory of supply chains with "transit times" to tree-structure networks, and present an exact treatment, which is uniformly applicable to various material flow topologies such as serial, assembly, and distribution systems. This is done by utilizing the *backward method* proposed by Zhao and Simchi-Levi (2005), which characterized the performance of assemble-to-order (ATO) systems in various settings.

We first derive the recursive equations for the backorder delays at all stages in the tree-structure supply chains (§3). This allows us to characterize the dependencies of backorder delays across different stages, including their intricate correlations in the multistage assembly systems. Based on the recursive equations, we then discuss (1) the impact of safety stock positioning in serial, assembly, and distribution systems and (2) the impact of the correlated delays on multistage assembly systems. We next numerically quantify the impact of the correlated delays in assembly systems and develop insights into the conditions under which the correlations may or may not be ignored (§4). Using the expected backorder delays as the decision variables, we modify Graves and Willems's (2000) algorithm based on a two-moment approximation to coordinate the base-stock policies in supply chains with stochastic sequential lead times (§5). Finally, through numerical examples (§6), we demonstrate the effectiveness of the approach, the impact of stock positioning in serial systems, and the impact of lead-time uncertainties.

It is perhaps worth mentioning that Assumption 1.1 ignores production capacity constraints. Thus, the framework developed in this paper can best be applied only to supply chains with excessive production capacities. If production capacities are the major concerns, then coordinating inventory policies in multistage supply chains poses a substantial challenge (van Houtum et al. 1996). Assuming that each

stage controls its inventory by a modified base-stock policy, Glasserman and Tayur (1995) developed a simulation-based optimization algorithm in capacitated serial supply chains. Kapuscinski and Tayur (1999) extended the methodology to handle more general networks, and Glasserman and Tayur (1996) developed simple and fast approximations.

2. The Model

Following Lee and Billington (1993), a node (a facility or a stage, equivalently) in the supply chain consists of a processor and a stock-keeping location. The links between every pair of nodes that have a supply and demand relationship are the lead time (from the supplier to the customer) and the demand process (from the customer to the supplier). Each node uses a continuous-time base-stock policy to manage its inventory. We consider supply chains where each node manages a single final item, one unit of which is assembled from one unit of possibly multiple different components. Note that the model can be extended to handle cases where multiple units of certain components are required to assemble one unit of the final item. For these components, we need to redefine the basic flow unit to be the sum of all the units required.

There are three types of lead times associated with each node in the network: (1) the lead times from the immediate suppliers, (2) the processing cycle time at the current node, and (3) backorder delays (or stockout delay, equivalently) to the immediate customers. The backorder delays are the delays due to stockouts at the current node. The lead time from a particular supplier is the sum of the transportation lead time from that supplier and its backorder delay. Finally, the processing cycle time is the time interval starting from an item being released into the processor until the time it completes processing.

The demand process at each node can be determined by aggregating the bill of materials. Because external demand follows independent Poisson processes, the demand at each node of the network with a tree structure also follows a Poisson process. Demand is satisfied on a first-come, first-served (FCFS) basis at all nodes, and unsatisfied demand at all nodes is fully backlogged. We assume that the probability distributions of the lead times from the external suppliers (suppliers outside of the supply chain under consideration) are known. Finally, the service level required by the external customers can be specified by a certain fill rate within a certain committed service time.

The following parameters are identified for the system:

- *Characteristics of the supply chain infrastructure*: the probability distributions of the processing cycle times and the transportation lead times as well as the replenishment lead times from the external suppliers; the inventory holding cost at each node.
- *Characteristics of the product and demand*: bill of materials, external demand processes, and service requirements of the external customers.

The supply chain is modeled by a tree graph $(\mathcal{N}, \mathcal{A})$ with the node set $\mathcal{N}$ and edge set $\mathcal{A}$. We index the nodes in $\mathcal{N}$ from $1, \ldots, K$; these nodes represent supply chain stages. An arc in $\mathcal{A}$, denoted by $(i, k) \in \mathcal{A}$, represents a pair of nodes in $\mathcal{N}$ that have the supply and demand relationship. For each node $k \in \mathcal{N}$, let n_k be a forward index of demand arrivals at node k. Because our focus is on systems in infinite time horizon, $n_k \in (-\infty, \ldots, -1, 0, 1, \ldots, \infty)$. We use the following notation:

- $X_k(n_k) =$ the backorder delay at node k for the demand arrival n_k.
- $W_k(n_k) =$ the inventory holding time of the corresponding item at stage k that satisfies the demand arrival n_k.
- $L_k(n_k) =$ the total replenishment lead time at node k for the order triggered by the demand arrival n_k.
- $P_k =$ the processing cycle time at node k.
- $t_{i,k} =$ the transportation lead time from node i to k, $(i, k) \in \mathcal{A}$.
- $S_k =$ the maximum of the lead times from external suppliers of node k. If node k does not have an external supplier, $S_k = 0$.
- $h_k =$ inventory holding cost per item per unit of time at node k.
- $s_k =$ base-stock level at node k.
- $\lambda_k =$ demand rate at node k.

By Assumption 1.1, P_k, $t_{i,k}$, and S_k are statistically the same for all n_k. Thus we suppress their dependence on n_k to simplify the notation.

Throughout this paper, we use the following notational conventions: $a^+ = \max\{a, 0\}$ and $E(\cdot)$, $V(\cdot)$, and

$\sigma(\cdot)$ are the mean, variance, and standard deviation of a random variable, respectively.

3. The Framework

In this section, we derive the recursive equations for the backorder delays in supply chains with a tree structure. We first focus on the general point demand processes, and then discuss systems facing independent Poisson processes as a special case. Based on the recursive equations, we develop analytical insights into safety stock positioning in various supply chain topologies.

Note that the interarrival time between two consecutive demand arrivals can be zero for point demand processes. In this case, each demand arrival index n_k is associated with a unit demand and represents the sequence in which the unit demands are satisfied at node k. For each node $k \in \mathcal{N}$, we first have the following definition:

DEFINITION 3.1. *$T_k(s_k, n_k)$ is the total time if one starts at the arrival time of demand n_k, counts backward until the number of demand arrivals at stage k reaches s_k.*

In other words, $T_k(s_k, n_k)$ is the sum of the interarrival times at node k between the demand arrivals n_k and n_k-1, n_k-1, and $n_k-2, \ldots,$ up to n_k-s_k+1 and n_k-s_k. Note that $T_k(s_k, n_k)$ depends on three factors: (1) the demand process at node k, (2) the base-stock level s_k, and (3) the demand arrival n_k.

We assume that an order can be split, i.e., whenever demand exceeds the inventory on hand, every stage delivers as much as possible from its on-hand inventory and backorders the rest (a similar assumption is also made by Zipkin 1991). Now, suppose that the demand n_k arrives at node k at time t; then, according to Zhao and Simchi-Levi (2005), the corresponding order of the node k that satisfies this demand is triggered by the demand arrival $n_k - s_k$, which arrives at node k at time $t_1 = t - T_k(s_k, n_k)$. This is true because of the FCFS assumption and the continuous-time base-stock policy (see Zhao and Simchi-Levi 2005 for more discussions). By the definition of L_k, the total replenishment lead time of the corresponding order is $L_k(n_k - s_k)$, and therefore it will arrive at node k at time $t_1 + L_k(n_k - s_k)$. Hence

$$X_k(n_k) = (L_k(n_k - s_k) - T_k(s_k, n_k))^+, \tag{1}$$

$$W_k(n_k) = (T_k(s_k, n_k) - L_k(n_k - s_k))^+. \tag{2}$$

Clearly, $L_k(n_k - s_k)$ depends on the supply network of node k, while $T_k(s_k, n_k)$ depends on the demand process at node k, and therefore the downstream network structure.

Let node i be an immediate supplier of node k. Clearly, the order triggered by the demand arrival $n_k - s_k$ at node k is a demand at node i. We denote the index of this demand at node i by $n_i(n_k)$ to indicate its dependence on the demand n_k at node k. Note that we suppress the dependence of $n_i(n_k)$ on s_k to simplify the notation. $n_i(n_k)$ can be a random variable because node i can face demand streams from nodes other than k. However, if node k is the only customer of node i, then the indices can be chosen so that $n_i(n_k) = n_k - s_k$. This is true because node i and k face the identical demand stream.

If node i has an immediate supplier node j, then following the same logic, the order triggered by the demand arrival $n_i(n_k) - s_i$ at node i is a demand at node j, whose index is $n_j(n_i(n_k))$. Because the supply chain has a tree structure, we can simplify the notation $n_j(n_i(n_k))$ by $n_j(n_k)$. In general, we use the notation $n_j(n_k)$ to denote the demand arrival at any upstream node j of node k that corresponds to the demand arrival n_k at node k. See Figure 1 for an example. If node k receives no supplies from internal nodes, then it follows from Assumption 1.1 that the $L_k(n_k - s_k)$ is statistically the same for all demand arrivals, and therefore we simplify the notation by L_k.

The key idea of this approach is that for each external demand arrival, we identify the corresponding order placed at each stage of the supply chain that eventually satisfies this demand. As we will see, this approach provides a uniform treatment to various supply chain topologies with stochastic sequential lead times.

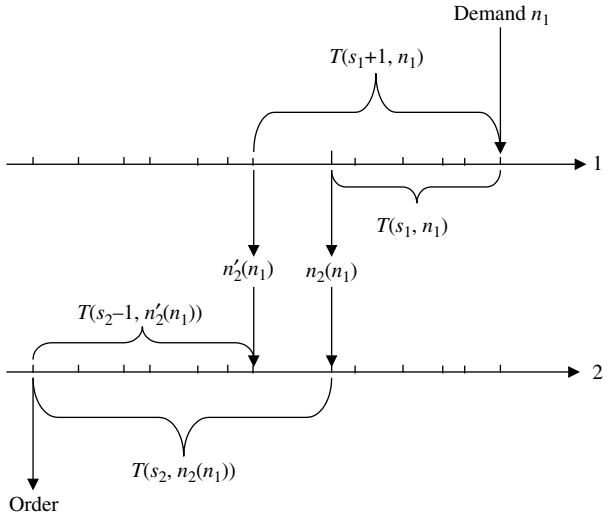
3.1. Serial Systems

Consider a serial supply chain with node k supplying node $k-1$ for $k = 2, 3, \ldots, K$, where node 1 faces external demand and node K receives external supplies. Because the demand processes are the same for all nodes in the supply chain, we drop the subscript k from T_k.

It follows from Equations (1)–(2) that the backorder delay of node 1 for demand arrival n_1 can be characterized by

$$X_1(n_1) = (L_1(n_1 - s_1) - T(s_1, n_1))^+, \tag{3}$$

Figure 1 **The Demand Arrivals and Their Corresponding Orders in a Serial System**



$$W_1(n_1) = (T(s_1, n_1) - L_1(n_1 - s_1))^+, \tag{4}$$

where

$$L_1(n_1 - s_1) = X_2(n_2(n_1)) + t_{2,1} + P_1. \tag{5}$$

Recall that for $k = 2, \ldots, K-1$, the demand arrival at node k that corresponds to the external demand arrival n_1 is denoted by $n_k(n_1)$. Therefore,

$$X_k(n_k(n_1)) = (L_k(n_k(n_1) - s_k) - T(s_k, n_k(n_1)))^+, \tag{6}$$

$$W_k(n_k(n_1)) = (T(s_k, n_k(n_1)) - L_k(n_k(n_1) - s_k))^+, \tag{7}$$

$$L_k(n_k(n_1) - s_k) = X_{k+1}(n_{k+1}(n_1)) + t_{k+1,k} + P_k, \tag{8}$$

and

$$X_K(n_K(n_1)) = (L_K - T(s_K, n_K(n_1)))^+, \tag{9}$$

$$W_K(n_K(n_1)) = (T(s_K, n_K(n_1)) - L_K)^+, \tag{10}$$

where $L_K = S_K + P_K$ is stochastic, sequential, and exogenously determined with known probability distribution.

For convenience, denote $n_1(n_1) = n_1$. For serial systems, we can choose the arrival indices at all nodes in a way such that $n_k(n_1) = n_1 - s_1 - \cdots - s_{k-1}$, $k = 2, \ldots, K$. Then, it is easy to see that $T(s_k, n_k(n_1))$, $k = 1, 2, \ldots, K$ are not overlapping. Figure 1 provides a visual aid. If external demand follows a renewal process (of which Poisson is a special case), then $T(s_k, n_k(n_1))$, $k = 1, 2, \ldots, K$ are mutually independent. Combining this fact with Assumption 1.1, we can characterize the backorder delay for any demand at node k by characterizing $X_K, X_{K-1}, \ldots, X_{k+1}$ sequentially. In this way, the serial supply chain can be decomposed into K single-stage systems each with independent replenishment lead time from the immediate upstream stage. It is perhaps worth mentioning that for renewal demand processes, we can suppress the dependence of X_k, L_k, and T on the index n_k, and therefore the recursive Equations (3), (5), (6), (8), and (9) are identical to Equation (8.3.4) of Zipkin (2000, p. 304), except that we focus on the backorder delays instead of the backorders.

We now prove the following general system property. Let $\leq_{\text{st}}$ stand for stochastic ordering.

THEOREM 3.2. *Consider two nodes, without loss of generality, node* 1 *and* 2, *in a supply chain with a tree structure under Assumption* 1.1 *and the assumption that orders can be split. Let node* 2 *be the only supplier of node* 1, *and let node* 1 *be the only customer of node* 2. *Node* 1 *faces a point demand process. Suppose* $s_2 > 0$, *and consider two stock positions* (*referred to as SP and SP′*) *that satisfy* $s_2' = s_2 - 1$ *and* $s_1' = s_1 + 1$. *Define the system with SP and the system with SP′ in the same probability space, i.e., the same demand streams and the same realizations of lead times. Then, for any demand arrival* n_1,

$$X_1'(n_1) \leq_{\text{st}} X_1(n_1). \tag{11}$$

For the item that satisfies the demand n_1 *at stage* 1,

$$W_2'(n_2'(n_1)) + W_1'(n_1) \leq_{\text{st}} W_2(n_2(n_1)) + W_1(n_1). \tag{12}$$

PROOF. See Appendix A for a proof. □

Intuitively, this theorem implies that moving one unit of stock downstream in any two serially linked nodes results in stochastically shorter backorder delays at the downstream node and stochastically shorter total waiting times for any item traveling through both nodes. This theorem is general because it applies to any tree-structure supply chain facing any point demand processes under mild assumptions.

Gallego and Zipkin (1999) pointed out that if the inventory holding costs are the same across all stages in the serial systems, safety stock should be located only at the last stage that faces external demand. We provide a stronger result as follows: consider a

K-stage serial supply chain. We refer to a stock positioning as SP if the base-stock levels satisfy $s_k > 0$ for some $k > 1$. We refer to a stock positioning as SP′ if the base-stock levels satisfy $s'_k = 0$, $\forall k \neq 1$, and $s'_1 > 0$. The following proposition is based on Theorem 3.2.

Proposition 3.3. *Consider the K-stage serial supply chain under Assumption* 1.1 *and the assumption that orders can be split. Let the two stock positions SP and SP′ satisfy* $s'_1 = \sum_{k=1}^{K} s_k$. *Assume that external demand follows a point process, and the systems with SP and SP′ are defined in the same probability space. Then, for any demand arrival* n_1,

$$X'_1(n_1) \leq_{\text{st}} X_1(n_1), \tag{13}$$

and for the item that satisfies this demand at stage 1,

$$W'_1(n_1) \leq_{\text{st}} \sum_{k=1}^{K} W_k(n_k(n_1)). \tag{14}$$

Proof. See Appendix A for a proof. □

One direct corollary of Proposition 3.3 is that safety stock should be positioned only at stage 1 if the inventory holding costs are the same across all stages. This result is valid in both the penalty cost model and the fill-rate constraint model (see Boyaci and Gallego 2001 for a definition of these models). The intuition, as illustrated by Theorem 3.2 and Proposition 3.3, is that products can move throughout the supply chain faster if inventory is kept only at the most downstream node. Thus, holding inventory only at stage 1 (stochastically) reduces the total product waiting times in the system as well as the backorder delays to the external customers. The magnitude of the impact of safety-stock positioning in serial supply chains is quantified in §6.2.

3.2. Assembly Systems

For simplicity, we assume that node k only has internal suppliers, $i = 1, 2, \ldots, I$. The case in which node k also has external suppliers is discussed at the end of §3. We make the following assumption as in Song and Zipkin (2002): when a demand arrives and some of its required components are in stock but others are not, we put the in-stock components aside as "committed stock."

Suppose that the demand n_k arrives at stage k at time t. Then the corresponding orders of components that satisfy this demand are triggered by the demand arrival $n_k - s_k$, and they are received by the suppliers $i = 1, 2, \ldots, I$ at the same time $t_1 = t - T_k(s_k, n_k)$ with index $n_i(n_k)$. It follows from Equation (1) that the backorder delay of the demand arrival $n_i(n_k)$ at node i is $X_i(n_i(n_k)) = (L_i(n_i(n_k) - s_i) - T_i(s_i, n_i(n_k)))^+$, and the total replenishment lead time at node k is equal to

$$\begin{aligned} &L_k(n_k - s_k) \\ &= \max_{i=1,2,\ldots,I} \{X_i(n_i(n_k)) + t_{i,k}\} + P_k \\ &= \max_{i=1,2,\ldots,I} \{(L_i(n_i(n_k) - s_i) - T_i(s_i, n_i(n_k)))^+ + t_{i,k}\} + P_k. \end{aligned} \tag{15}$$

Clearly, assembly systems are more complicated than the serial systems, not only because the total replenishment lead time is determined by the maximum of the lead times from all suppliers, but also because the backorder delays of the suppliers, $X_i(n_i(n_k))$, $i \in I$, are correlated. To see this, we analyze a simple case in which node k is the only customer of all its suppliers. Thus $n_i(n_k) = n_k - s_k$, $\forall i = 1, 2, \ldots, I$, and we drop subscript i accordingly from T_i because all suppliers face the same demand process. We index the suppliers in a nondecreasing order of their base-stock levels; then according to Zhao and Simchi-Levi (2005),

$$T(s_i, n_k - s_k) = T(s_{i-1}, n_k - s_k) + T(s_i - s_{i-1}, n_k - s_k - s_{i-1}), \quad i = 2, 3, \ldots, I. \tag{16}$$

To show that this equation is true, we consider two suppliers i and i' of node k. If $s_i \leq s_{i'}$, then by Definition 3.1, $T(s_{i'}, n_k - s_k)$ overlaps with $T(s_i, n_k - s_k)$ for the time period $[t_1 - T(s_i, n_k - s_k), t_1]$.

If some suppliers of node k also serve other assembly systems, the analysis becomes much more complicated because one assembly system can order from more than one supplier and one supplier may serve one or more assembly systems. The material flow topology of such a system is similar to that of a multicomponent and multiproduct ATO system in which each product may request more than one component and one component may be requested by more than one product. We refer the reader to Zhao and Simchi-Levi (2005) for more discussion of the multicomponent and multiproduct ATO systems. In what follows, we assume that each assembly node is the only customer of its suppliers.

Stock positioning has an impact on the backorder delays in assembly systems.

PROPOSITION 3.4. *Consider the assembly node k under Assumption* 1.1 *and the assumption that orders can be split. Assume that node k is the only customer of its suppliers $i = 1, 2, \ldots, I$ and it faces a point demand process. Suppose $s_i > 0$, $\forall i$, and consider two stock positions (referred to as SP and SP′) that satisfy $s_i' = s_i - 1$, $\forall i = 1, 2, \ldots, I$, and $s_k' = s_k + 1$. Define the systems with SP and SP′ in the same probability space, then, for any demand arrival n_k,*

$$X_k'(n_k) \leq_{\text{st}} X_k(n_k). \tag{17}$$

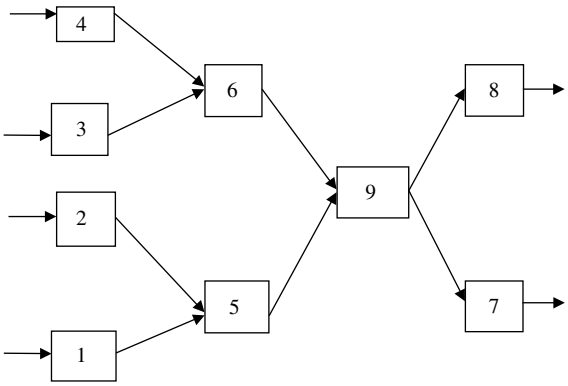
PROOF. See Appendix A for a proof. □

This proposition implies that reducing one unit of stock for all the components, while at the same time increasing one unit of stock for the finished item at the assembly node, results in stochastically shorter backorder delays at the assembly node. However, it is not clear how to characterize the impact of stock positioning on the total waiting time of a finished item at node k and its components at node k and nodes $i = 1, 2, \ldots, I$, because of the extra time that a component has to wait at node k for other corresponding components to be replenished.

We next consider multistage assembly systems. Suppose each supplier i is, in turn, supplied by other node(s) in the system. Then $L_i(n_i(n_k) - s_i)$ in Equation (15) may be correlated across nodes $i = 1, 2, \ldots, I$. In addition, if $s_i < s_{i'}$, then $L_i(n_i(n_k) - s_i)$ may be correlated with $T(s_{i'}, n_{i'}(n_k))$.

To demonstrate the intricate correlations in the multistage assembly systems, we consider the example illustrated in Figure 2, which is a nine-stage production-distribution system with stochastic processing times. For ease of exposition, we assume zero transportation lead times among internal nodes and zero replenishment lead times from the external suppliers. Consider an arbitrary demand arrival n_9 at node 9, and let its arrival time be t. We index all the demand interarrival times prior to time t by ν_n, $n = 1, 2, \ldots,$ where ν_n is the nth interarrival time if we start at time t and count *backward*. Applying Equations (1) and (15), we can characterize $X_9(n_9)$ as follows:

Figure 2 A Multistage Production-Distribution System



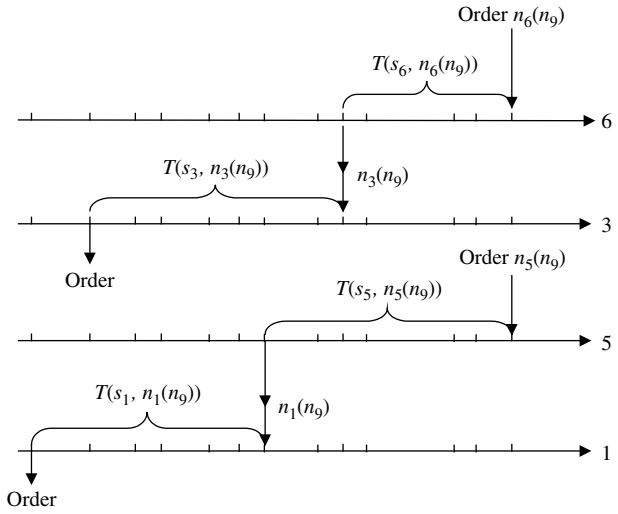
$$\begin{aligned}
X_9(n_9) &= (L_9(n_9 - s_9) - T(s_9, n_9))^+, \\
T(s_9, n_9) &= \sum_{n=1}^{s_9} \nu_n, \qquad (18)\\
L_9(n_9 - s_9) &= \max\{X_5(n_5(n_9)), X_6(n_6(n_9))\} + P_9, \\
X_5(n_5(n_9)) &= (L_5(n_5(n_9) - s_5) - T(s_5, n_5(n_9)))^+, \\
T(s_5, n_5(n_9)) &= \sum_{n=s_9+1}^{s_9+s_5} \nu_n, \qquad (19)\\
L_5(n_5(n_9) - s_5) &= \max\{X_1(n_1(n_9)), X_2(n_2(n_9))\} + P_5, \\
X_1(n_1(n_9)) &= (P_1 - T(s_1, n_1(n_9)))^+, \\
T(s_1, n_1(n_9)) &= \sum_{n=s_9+s_5+1}^{s_9+s_5+s_1} \nu_n, \\
X_2(n_2(n_9)) &= (P_2 - T(s_2, n_2(n_9)))^+, \\
T(s_2, n_2(n_9)) &= \sum_{n=s_9+s_5+1}^{s_9+s_5+s_2} \nu_n, \qquad (20)
\end{aligned}$$

and

$$\begin{aligned}
X_6(n_6(n_9)) &= (L_6(n_6(n_9) - s_6) - T(s_6, n_6(n_9)))^+, \\
T(s_6, n_6(n_9)) &= \sum_{n=s_9+1}^{s_9+s_6} \nu_n, \qquad (21)\\
L_6(n_6(n_9) - s_6) &= \max\{X_3(n_3(n_9)), X_4(n_4(n_9))\} + P_6, \\
X_3(n_3(n_9)) &= (P_3 - T(s_3, n_3(n_9)))^+, \\
T(s_3, n_3(n_9)) &= \sum_{n=s_9+s_6+1}^{s_9+s_6+s_3} \nu_n, \\
X_4(n_4(n_9)) &= (P_4 - T(s_4, n_4(n_9)))^+, \\
T(s_4, n_4(n_9)) &= \sum_{n=s_9+s_6+1}^{s_9+s_6+s_4} \nu_n. \qquad (22)
\end{aligned}$$

Figure 3 provides a visual aid. It is easily seen that $T(s_3, n_3(n_9))$ can be correlated with $T(s_1, n_1(n_9))$,

Figure 3 The Correlation Structure of Lead Times in a Multistage Assembly System



and therefore it follows from Equations (19)–(22) that $L_6(n_6(n_9) - s_6)$ can be correlated with $L_5(n_5(n_9) - s_5)$. If $s_5 > s_6$, then $T(s_3, n_3(n_9))$ is correlated with $T(s_5, n_5(n_9))$, which implies that $L_6(n_6(n_9) - s_6)$ is correlated with $T(s_5, n_5(n_9))$. To summarize, the backorder delays of parallel branches in a multistage assembly system can be correlated. If external demand follows independent Poisson processes, we can characterize the joint distribution of the backorder delays at all nodes by identifying the right overlapping interarrival times.

To analyze the impact of correlated lead times on system performance, we need the concept of the *associated* random variables. Consider random variables $Y_1, Y_2, \ldots, Y_n$, and the vector $\overline{Y} = (Y_1, Y_2, \ldots, Y_n)$. The following definition is due to Esary et al. (1967).

Definition 3.5. *The random variables $Y_1, Y_2, \ldots, Y_n$ are associated if*

$$Cov[\phi(\overline{Y}), \psi(\overline{Y})] \geq 0, \tag{23}$$

or equivalently

$$E[\phi(\overline{Y})\psi(\overline{Y})] \geq E[\phi(\overline{Y})]E[\psi(\overline{Y})] \tag{24}$$

for all nondecreasing real functions ϕ, ψ for which $E[\phi(\overline{Y})]$, $E[\psi(\overline{Y})]$, and $E[\phi(\overline{Y})\psi(\overline{Y})]$ exist.

We refer to Tong (1980) and Shaked and Shanthikumar (1994) for reviews. Associated random variables have the following important property (see, e.g., Esary et al. 1967, Theorem 5.1).

Property 3.6. *Let $Y_1, Y_2, \ldots, Y_n$ be associated random variables; then*

$$\Pr\{Y_1 \leq y_1, Y_2 \leq y_2, \ldots, Y_n \leq y_n\} \geq \prod_{k=1}^{n} \Pr\{Y_k \leq y_k\}$$

and

$$\Pr\{Y_1 > y_1, Y_2 > y_2, \ldots, Y_n > y_n\} \geq \prod_{k=1}^{n} \Pr\{Y_k > y_k\}$$

for all $y_1, y_2, \ldots, y_n$.

Proposition 3.7 characterizes the impact of the dependent lead times on the assembly systems. The proof involves establishing the association of the backorder delays.

Proposition 3.7. *Consider an assembly node k and its suppliers $i = 1, 2, \ldots, I$ in a tree-structure supply chain under Assumption* 1.1. *If external demand follows independent Poisson processes, then for any demand arrival n_k and any $\tau \geq 0$,*

$$\Pr\Big\{\max_{i=1,2,\ldots,I}\{X_i(n_i(n_k)) + t_{i,k}\} \leq \tau\Big\} \geq \prod_{i=1,2,\ldots,I} \Pr\{X_i(n_i(n_k)) + t_{i,k} \leq \tau\}. \tag{25}$$

Proof. See Appendix A for a proof. □

The intuition behind this proposition is that the backorder delays, $X_i(n_i(n_k))$, are driven by the common demand interarrival times; thereby they tend to "hang on" together. Note that this proposition does not require the assumption that each assembly node is the only customer of its suppliers. Combining this proposition with Equations (15), (1), and (2), we conclude that a multistage assembly system with dependent lead times has stochastically shorter delay, $X_k(n_k)$, and longer waiting time, $W_k(n_k)$, at each assembly node k than the delay and waiting time in a corresponding system where the dependence of the lead times are ignored.

3.3. Distribution Systems

Consider a distribution node k with customer nodes $j = 1, 2, \ldots, J$. For the demand arrival n_j at node j, it follows from the same logic as in serial systems that

$$X_j(n_j) = (L_j(n_j - s_j) - T_j(s_j, n_j))^+, \tag{26}$$

$$W_j(n_j) = (T_j(s_j, n_j) - L_j(n_j - s_j))^+, \tag{27}$$

where

$$L_j(n_j - s_j) = X_k(n_k(n_j)) + t_{k,j} + P_j, \quad (28)$$

$$X_k(n_k(n_j)) = (L_k(n_k(n_j) - s_k) - T_k(s_k, n_k(n_j)))^+, \quad (29)$$

$$W_k(n_k(n_j)) = (T_k(s_k, n_k(n_j)) - L_k(n_k(n_j) - s_k))^+, \quad (30)$$

where $L_k(n_k(n_j) - s_k)$ depends on the supply network of node k.

Note that $T_k(s_k, n_k(n_j))$ is based on the superposition of the demand processes of all customer nodes, while $T_j(s_j, n_j)$ is based on the demand process of customer node j only. If external demand follows independent Poisson processes, then it follows from the fact that $T_j(s_j, n_j)$ does not overlap with $T_k(s_k, n_k(n_j))$, that $T_j(s_j, n_j)$ is independent of $T_k(s_k, n_k(n_j))$. In addition, $T_k(s_k, n_k)$ is statistically the same for every $n_k \in \{-\infty, \ldots, -1, 0, 1, \ldots, \infty\}$ because the superposition of independent Poisson processes is still a Poisson process.

In practice, most companies locate their warehouses closer to retail outlets than to manufacturing facilities. Proposition 3.8 provides a justification for this phenomenon from the inventory perspective.

Proposition 3.8. *Consider a supply chain in which node i is the only supplier of node k, and node k is the only supplier of nodes $j = 1, 2, \ldots, J$. Each node j faces a point demand process, which may be correlated across $j = 1, 2, \ldots, J$. Assumption* 1.1 *holds and orders can be split. Assume deterministic transportation lead times, and consider two locations of node k that satisfy*

$$t_{i,k} + t_{k,j} = t'_{i,k} + t'_{k,j} \quad (31)$$

and

$$t'_{k,j} < t_{k,j} \quad (32)$$

for a certain node j. Define the supply chain with different locations in the same probability space, and keep the base-stock levels at all nodes the same for both locations. Then, for any demand arrival n_j at node j,

$$X'_j(n_j) \leq_{\text{st}} X_j(n_j), \quad (33)$$

and for the item that satisfies this demand at node j,

$$W'_k(n_k(n_j)) + W'_j(n_j) \leq_{\text{st}} W_k(n_k(n_j)) + W_j(n_j). \quad (34)$$

Proof. See Appendix A for a proof. □

Intuitively, Proposition 3.8 shows that keeping the stock levels at all facilities the same and keeping the total transportation lead times between the manufacturer and retail outlets unchanged, then the closer the distribution facility to the retail outlets, the (stochastically) shorter the backorder delays to external customers and the (stochastically) shorter the time any item stays in the system.

To study the impact of stock positioning in distribution systems, we analyze a simple case in which the customer nodes are identical; i.e., they face identical demand processes, and have identical transportation lead times $t_{k,j}$ and base-stock levels s_j for all $j = 1, 2, \ldots, J$. Assuming $s_j > 0, \forall j$, we consider two stock positions (referred to as SP and SP′) that satisfy $s'_j = s_j - 1, \forall j = 1, 2, \ldots, J$, and $s'_k = s_k + J$, that is, moving one unit of stock from each customer node to the distribution node. In this system, however, stochastic inequalities similar to those established in Theorem 3.2 do not hold because of the risk-pooling effect. Indeed, a numerical study (not reported here) shows that moving stock upstream can reduce the expected backorder delays to the external customers.

3.4. Systemwide Properties

Proposition 3.9 characterizes an important property for every node in the supply chain.

Proposition 3.9. *Under Assumption* 1.1 *and the assumption of independent Poisson demand processes, $L_k(n_k - s_k)$ is independent of $T_k(s_k, n_k)$ for any demand arrival n_k at every node $k \in \mathcal{N}$.*

Proof. Consider an arbitrary node $k \in \mathcal{N}$. Suppose a demand arrival n_k is realized at time t. Then the corresponding order is placed at time $t - T_k(s_k, n_k)$. Furthermore, the total replenishment lead time at node k, $L_k(n_k - s_k)$, for this specific order is determined by the demand interarrival times prior to $t - T_k(s_k, n_k)$ as well as the processing cycle times, transportation lead times, and replenishment lead times from external suppliers. The assumption of the independent Poisson processes implies that nonoverlapping interarrival times are independent. Therefore it follows from Assumption 1.1 that $L_k(n_k - s_k)$ is independent of the $T_k(s_k, n_k)$ for every n_k at every $k \in \mathcal{N}$. □

Combining Proposition 3.9 with the recursive equations in §§3.1–3.3, it is possible to exactly characterize the joint probability distribution of the backorder delays in tree-structure supply chains with stochastic sequential lead times when the external demand follows independent Poisson processes and each assembly node in the network is the only customer of its suppliers. Observe that the backorder delays are determined by the "transit times" and the demand interarrival times; it follows from Assumption 1.1 and the assumption of independent Poisson processes that the L_k, T_k, X_k, and W_k are statistically the same for all demand arrivals n_k at each node $k \in \mathcal{N}$. This is true because the reversed Poisson processes are still Poisson and independent of the starting time (Kulkarni 1995). For the rest of this paper, we focus on independent Poisson demand processes, and thereby drop the index n_k from these random variables without causing any confusion.

Because of the correlations of the backorder delays in multistage assembly systems, an exact characterization (unless it is based on Monte-Carlo simulation) of the system performances poses a significant analytical and numerical challenge. To improve computational efficiencies, we present approximations in the next section. We also study the accuracies of the approximations by comparing them to a fast simulation based on the exact recursive equations.

To summarize, the backorder delay at each node, X, is fully determined by the total replenishment lead time, L, base-stock level, s, and the demand process. Conversely, given L, $E(X)$, and the demand process, s is also fully determined (because of the monotonic relationship between s and $E(X)$). Thus, given s_k or $E(X_k)$ at all nodes $k \in \mathcal{N}$, the steady-state performance of a supply chain with tree structure is completely determined. By Little's law, the steady-state average inventory level, $E(I_k)$, at node k for finish goods satisfies

$$\begin{aligned} E(I_k) &= \lambda_k \times E(W_k) \\ &= \lambda_k \times E((T_k(s_k) - L_k)^+) \\ &= \lambda_k \times (E(T_k(s_k)) - E(L_k) + E(X_k)). \end{aligned} \tag{35}$$

If an assembly node k has both internal and external suppliers, then

$$L_k = \max\{X_i + t_{i,k},\ \forall (i,k) \in \mathcal{A};\ S_k\} + P_k. \tag{36}$$

Observe that a component may need to stay at an assembly node without being processed before all other corresponding components arrive; thus it incurs inventory holding costs. To determine the steady-state average inventory level, $E(I_k^i)$, at an assembly node k for a component i, $\forall (i,k) \in \mathcal{A}$, it follows from Little's law and Equation (36) that

$$\begin{aligned} E(I_k^i) &= \lambda_k \times E(\max\{X_j + t_{j,k},\ \forall (j,k) \in \mathcal{A};\ S_k\} - X_i - t_{i,k}) \\ &= \lambda_k \times E(L_k - P_k - X_i - t_{i,k}). \end{aligned} \tag{37}$$

We only focus on the expected holding costs for inventories I_k and I_k^i at all nodes, because the expected inventories carried during the transportation lead times and processing cycle times are constants.

4. Approximations

The objective of this section is to develop and test tractable approximations. More numerical studies are conducted in §6 to test the accuracy of the approximations in various examples.

4.1. The Lead-Time Correlations

In this section, we quantify the impact of the correlated lead times on system performances, and develop insights into the conditions under which the correlations may or may not be ignored using simple but representative examples.

For this purpose, we consider an assembly node k and its suppliers $i = 1, 2, \ldots, I$. For simplicity, we assume that the total replenishment lead times at the suppliers, L_i, are mutually independent and also independent of the system state. We focus on the cases where node k is the only customer of its suppliers because the correlations among the suppliers' delays, X_i, $i = 1, 2, \ldots, I$, tend to be weaker if some of the suppliers also serve other customers. This is true because the correlation is due to the correlated demand processes at the suppliers (see, e.g., Equation (16)). If node k is the only customer of its suppliers, the demand processes at the suppliers are identical, and therefore completely correlated. On the other hand, if supplier i only serves node k but supplier i' also serves other nodes, then the demand processes at the suppliers i and i' are not identical, and thereby less correlated. By Equation (15), we have

$$L_k = \max_{i=1,2,\ldots,I} \{(L_i - T(s_i))^+ + t_{i,k}\} + P_k, \tag{38}$$

where s_i is the base-stock level at supplier i, $i = 1, 2, \ldots, I$, and we drop the subscript i from T_i because of the identical demand process. To highlight the impact of the correlations among $T(s_i)$, $i = 1, 2, \ldots, I$, we assume $P_k = 0$ and $t_{i,k} = 0$, $\forall i$.

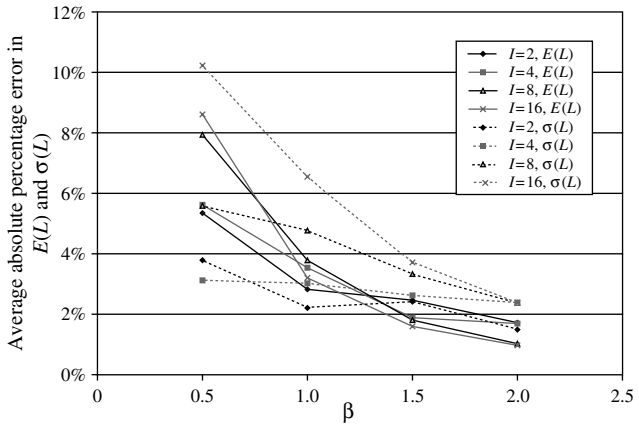
To quantify the impact of the correlations in this system, we focus on the absolute percentage errors in $E(L_k)$ and $\sigma(L_k)$ between systems with correlated lead times and systems with independent lead times. For instance, the absolute percentage errors in $E(L_k)$ is defined as the absolute difference between the $E(L_k)$'s in the two systems divided by $E(L_k)$ in the system with correlated lead times. To generate a sample of L_k with correlated lead times, we utilize Equation (16); to generate a sample of L_k with independent lead times, we sample $T(s_i)$ independently.

Without loss of generality, let $\lambda_k = 1$. We assume that L_i follows an Erlang distribution with parameters μ^i and n^i (see Zipkin 2000 for a definition) because it is a special case of both Gamma and continuous phase-type distributions, and is often used in practice to model lead times (Zipkin 2000). Thus the following nondimensional parameters may have an impact on the absolute percentage errors: I, n^i (where $\sigma(L_i)/E(L_i) = 1/\sqrt{n^i}$) and $\lambda_k E(L_i)/s_i$, $i = 1, 2, \ldots, I$. In the numerical study, we mainly focus on I and $E(L_i)/s_i$ (because $\lambda_k = 1$).

To construct a test instance, we first generate s_i randomly according to $\lceil Uniform(0, 10) \rceil$ distribution, and then generate $E(L_i)$ randomly according to $\beta \times Uniform(0, 10)$, where β is a scaling factor representing the average ratio between $E(L_i)$ and s_i across all i. Clearly, the greater the β, the higher the $E(X_i)/E(L_i)$, and the lower the fill rates at the suppliers. For a given n^i, we compute $\mu^i = n^i/E(L_i)$ and $\sigma(L_i) = E(L_i)/\sqrt{n^i}$. For any combination of the nondimensional parameters $\beta = 0.4, 0.8, 1.4, 2$, and $I = 2, 4, 8, 16$, we generate 100 sample systems by randomly choosing s_i and $E(L_i)$. For each sample system, we simulate 10,000 runs with or without considering the correlations, and calculate the absolute percentage error in $E(L_k)$ and $\sigma(L_k)$. Finally, we take the average of the absolute percentage errors over the 100 sample systems.

A numerical study shows that $E(L_k)$ increases as we switch from systems with correlated lead times to corresponding systems with independent lead times,

Figure 4 **The Impact of the Lead-Time Correlations on $E(L_k)$ and $\sigma(L_k)$**



which confirms Proposition 3.7. Figure 4 shows the case of $n^i = 5$, $\forall i$. The cases of $n^i = 1, 3, 7$ are also studied (not reported here). Summarizing all the numerical results, we observe that

- Ignoring the correlations may result in relatively large errors in both $E(L_k)$ and $\sigma(L_k)$ (about 10% error) when β is relatively small; e.g., $\beta = 0.4$, where the fill rates are around 80% in the numerical examples.
- When β is relatively large, e.g., $\beta = 2$, where the fill rates are around 30% in the numerical examples, the percentage errors may be relatively small, i.e., no more than 3%.

These observations are intuitive because as β increases, $E(L_i)$, $i \in I$ increase and it follows from equation $X_i = (L_i - T(s_i))^+$ that the correlation between X_i and X_j, $i, j = 1, 2, \ldots, I$ becomes weaker because L_i, $i = 1, 2, \ldots, I$ are mutually independent. On the other hand, there is no clear trend in the relationship between the percentage errors and the number of suppliers, I. That is, in some cases as the number of the suppliers increases, the error also increases; while in others, the error decreases.

Note that the numerical study is conducted at $P_k = 0$ and $t_{i,k} = 0$ for all i. It follows from Assumption 1.1 that when P_k is substantial with respect to T_i, the percentage errors of ignoring the correlations should be much smaller in all cases. Clearly, if $t_{i,k}$, $\forall i$ are also positive, then the percentage errors should be even smaller. The numerical study implies that the lead-time correlations may not be ignored in systems where the transportation and processing lead times are negligible relative to L_i, e.g., the ATO systems

(Song and Zipkin 2002). However, the errors become more tolerable as the transportation and processing lead times increase.

4.2. Two-Moment Approximations

Computing the probability distribution of the backorder delay at each stage of a supply chain is time consuming, especially for the purpose of system optimization. To further improve the computational efficiency, we utilize two-moment approximations, which have been studied extensively in inventory systems. We refer to Axsäter (1993) for a review of two-moment approximations in distribution systems.

We utilize the approach by Graves (1985) and Svoronos and Zipkin (1991). In particular, given the total replenishment lead time L ($E(L)$ and $V(L)$), demand rate λ, and base-stock level s, we compute the backorder delay X ($E(X)$ and $V(X)$) as follows: (1) Compute the first two moments of the lead-time demand by $E(L)$, $V(L)$, and the demand process at each stage. (2) Fit the lead-time demand distribution by a negative-binomial distribution that matches the first two moments. (3) Calculate the first two moments of the backorder at the base-stock level s. (4) Find the mean and variance of the backorder delay by the first two moments of the backorder and the demand process. We refer the reader to Svoronos and Zipkin (1991) for a thorough discussion.

Conversely, given $E(X)$, the total replenishment lead time L ($E(L)$ and $V(L)$), and demand rate λ, we can calculate s and $V(X)$ as follows: Steps (1) and (2) of the above procedure remain the same, (3) calculate s using the lead-time demand distribution and $E(X)$, (4) compute the variance of the backorder using the base-stock level, and (5) compute the variance of the backorder delays. For assembly nodes, we use the two-moment approximation by Clark (1961) to calculate the mean and variance of the maximum of independent random variables.

Define $\overline{X}_k$ to be a vector representing the backorder delays of the immediate suppliers of node k; i.e., $\overline{X}_k = \{X_i \mid (i,k) \in \mathcal{A}\}$. For any node $k \in \mathcal{N}$, given the following parameters: the mean and variance of $\overline{X}_k$ ($E(\overline{X}_k)$ and $V(\overline{X}_k)$), the average backorder delay $E(X_k)$, and demand rate λ_k, we can determine the mean and variance of L_k by Equation (36), the safety stock carrying costs at node k, denoted by $H_k(E(\overline{X}_k), V(\overline{X}_k), E(X_k))$, as well as the variance of the backorder delay, $V(X_k)$. In particular,

$$H_k(E(\overline{X}_k), V(\overline{X}_k), E(X_k)) = h_k E(I_k) + \sum_{(i,k)\in\mathcal{A}} h_i E(I_k^i), \quad (39)$$

where $E(I_k)$ and $E(I_k^i)$ are defined by Equations (35)–(37). For convenience, denote

$$V(X_k) = f_k(E(\overline{X}_k), V(\overline{X}_k), E(X_k)). \quad (40)$$

5. Optimization

By using $E(X_k)$, $k = 1, 2, \ldots, K$ as decision variables, we can formulate the following program:

$$\begin{aligned}
\mathbf{P} \quad \min \ & \sum_{k=1}^{K} H_k(E(\overline{X}_k), V(\overline{X}_k), E(X_k)) \\
\text{s.t.} \ & V(X_k) = f_k(E(\overline{X}_k), V(\overline{X}_k), E(X_k)), \quad \forall k \in \mathcal{N}, \\
& L_k = \max\{X_i + t_{i,k}, \forall (i,k) \in \mathcal{A}; S_k\} + P_k, \\
& \qquad\qquad \forall k \in \mathcal{N}, \\
& 0 \le E(X_k) \le \min\{Q_k, E(L_k)\}, \quad \forall k \in \mathcal{N}, \\
& V(X_k) \le \sigma_k^2, \quad \forall k \in \mathcal{N}, \\
& \Pr\{X_k \le \tau_k\} \ge \alpha_k, \\
& \qquad \text{for node } k \text{ serving external customers.}
\end{aligned}$$

We assume that the service requirements, τ_k and α_k, are specified by the external customers, and the distributions of S_k, $k \in \mathcal{N}$ are known.

Program **P** can be explained as follows: the first constraint specifies the variance of the backorder delay, the second constraint relates the total replenishment lead time at node k to the backorder delays of the immediate upstream nodes, as well as the lead times from the external suppliers. The third and fourth constraints restrict the feasible regions for $E(X)$ and $V(X)$, and the last constraint specifies the service requirements of the external customers. Q_k and σ_k^2 are the maximum allowable expected backorder delay and delay variance at node k, respectively. More specifically, they are determined by the minimum of the most tolerable expected delays and delay variances among the customers of node k.

If node k serves external demand, then X_k is subject to the service constraint, $\Pr\{X_k \le \tau_k\} \ge \alpha_k$. It follows

from Equation (1) that we can rewrite the service constraint as $\Pr\{L_k - T_k(s_k) \le \tau_k\} \ge \alpha_k$. We fit $L_k - T_k(s_k)$ by a normal random variable, and therefore the constraint becomes

$$E(L_k) - s_k/\lambda_k + z_{1-\alpha_k}\sqrt{V(L_k) + s_k/\lambda_k^2} \le \tau_k. \qquad (41)$$

Evidently, program **P** is not linear. Graves and Willems (2000) developed an algorithm based on dynamic programming to optimize safety stock positioning in supply chains with guaranteed service time (see also Graves and Willems 2003a). We generalize their algorithm so as to coordinate stock levels in supply chains with stochastic sequential lead times. The difference between their algorithm and the algorithm developed in this paper is that we focus on the mean and variance of the stochastic service times, while Graves and Willems (2000) focus on the guaranteed service times. We refer the reader to Appendix B for the algorithm, an example, and some implementation details.

6. Numerical Studies

The objective of this section is twofold: (1) testing the accuracy of the approximations and the quality of the solution found by the dynamic programming (DP) algorithm and (2) developing insights with respect to stock positioning and the impact of the lead-time uncertainties.

6.1. Effectiveness of the Approach

The first example is the multistage system in Figure 2 with only a subset of the nodes, i.e., $\mathcal{N} = \{1, 2, 5, 6, 9\}$. All other nodes are regarded either as external suppliers or external customers. For simplicity, we assume zero external lead times and zero transportation lead times, but nonzero processing cycle times at all nodes. Nonzero transportation lead times are considered in §6.3. The stochastic processing times follow Erlang distributions. Without loss of generality, let $\lambda_9 = 1$, $h_1 = h_2 = 1$, $h_5 = 2$, $h_6 = 1.5$, and $h_9 = 3$. Note that inventory holding costs are increasing as one moves downstream in the supply chain.

To test the accuracy of the approximations, we first used the DP algorithm to find a solution and then used the Monte-Carlo simulation, which is based on the exact recursive equations in §3 to evaluate the solution. In the simulation, we ran 10^4 independent replications for each problem instance and calculated the 95% confidence interval for the performance measures.

Two studies are conducted. First, we set the lead-time parameters as follows: $\{n^1, n^2, n^5, n^6, n^9\} = \{7, 9, 2, 2, 9\}$ and $\{E(P_1), E(P_2), E(P_5), E(P_6), E(P_9)\} = \{10, 3, 9, 3, 3\}$, and vary τ_9 and α_9. For each parameter set, we use the DP algorithm to determine the base-stock level at each node and then use simulation to estimate the performance of this solution. Specifically, column "Actual α_9 (simulation)" in Table 1 presents the simulated service level at node 9 to the downstream facilities. Column "Cost (approximation)" is the cost generated by DP, while "Cost (simulation)" is generated by simulation. In the last column of Table 1, the absolute percentage difference in cost is defined as the absolute difference between the simulated cost and the cost generated by DP divided by the simulated cost.

Second, we set $\alpha_9 = 95\%$ and $\tau_9 = 4$, and randomly generate 100 systems with different lead-time parameters, where $n^i \sim Uniform\{1, 2, \ldots, 10\}$ and $E(P_i) \sim$

Table 1 Comparison Between the Approximation and the Simulation for the DP Solutions

τ_9	Target α_9 (%)	Actual α_9 (simulation)	Difference in fill rates (%)	Cost (approximation)	Cost (simulation)	Absolute percentage difference in costs (%)
2	99	96.75% ± 0.35%	2.25	49.64	50.19 ± 0.36	1.11
2	95	92.79% ± 0.51%	2.21	37.91	38.99 ± 0.32	2.77
2	90	90.19% ± 0.58%	0.19	32.57	33.63 ± 0.29	3.14
2	80	83.54% ± 0.73%	3.54	27.47	28.19 ± 0.26	2.55
0	95	92.83% ± 0.51%	2.17	44.14	45.11 ± 0.34	2.15
4	95	92.50% ± 0.52%	2.5	32.57	33.62 ± 0.29	3.13
6	95	92.79% ± 0.51%	2.21	26.13	26.73 ± 0.26	2.27
8	95	92.04% ± 0.53%	2.96	21.85	22.33 ± 0.23	2.12

Uniform(1, 10). Numerical results show that the average absolute percentage difference (between the simulation and the approximation) in cost is 1.79%, with a maximum of 5.86%; while the average absolute difference between the target α_9 and the simulated α_9 of the DP solution is 1.76%, with a maximum of 5.02%.

Given the fact that we only consider the first two moments of the lead times, as well as the other approximations detailed in §§4 and 5, the numerical study implies that the approximations are reasonably accurate for a fairly wide range of parameters. Additional studies (not reported here) show that the percentage errors are even smaller when the transportation lead times are nonzero.

A closer examination of the 100 randomly generated systems reveals that the least accurate cases have the smallest $E(P_9)$ over the average processing time of all nodes as well as the smallest n^i for all i. This is consistent with the observations made in §4: The smaller the $E(P_9)$ over the average processing time, the larger the impact of the errors in backorder delays in the last stage, and therefore the approximations may not perform well. To explain the impact of n^i on the accuracy of the approximations, we note that the larger the n^i, the more accurately the normal distribution fits the processing times. Conversely, when n^i is small, normal distribution is not a good fit, and when $n^i = 1$, the processing times follow exponential distributions. Thus the normal approximation of the fill-rate constraints and Clark's (1961) two-moment approximation, which applies to normal random variables, may yield sizeable errors for small n^i. Finally, we observe that the simulated fill rates of the DP solution may not closely match with the targets. To generate near optimal and feasible solutions, we can adjust the input fill rates and run the DP algorithm repetitively until the target fill rates fall into the 95% confidence interval of the simulated fill rates.

To test whether the DP solution is indeed optimal or close to optimal, we continue with the same example and compare the DP solution to a solution found by a simulation-based search algorithm. In the search algorithm, we first identify an upper bound $\tilde{s}_k$ for the base-stock level at each node. Then, for any base-stock level vector $\{s_1, s_2, s_5, s_6\} \in \bigotimes_{k=1,2,5,6}\{0, \lfloor \tilde{s}_k/10 \rfloor, \lfloor 2\tilde{s}_k/10 \rfloor, \ldots, \tilde{s}_k\}$, we first choose s_9 so that the simulated fill rate closely matches with the target, and then evaluate average system cost by simulation. For comparison, we adjust the input fill rates so that the simulated fill rates of the DP solutions closely match with the target. Table 2 summarizes the results for the cases of $\tau_9 = 4$ and $\alpha_9 = 90\%$, where the $\{E(P_1), E(P_2), E(P_5), E(P_6)\} = \{5, 5, 2, 6\}$, but $E(P_9)$ and $\{n^1, n^2, n^5, n^6, n^9\}$ are allowed to vary. All costs are evaluated by simulation.

The percentage difference in costs is defined as the difference between the cost of the DP solution and the cost of the search-based solution divided by the cost of the search-based solution. On a Pentium 1.67 GHz laptop, the search algorithm takes more than an hour for one instance, while the DP algorithm takes about one second.

We first note that in all cases, the cost of the DP solution is reasonably close to that of the search-based solution. We also observe that when n^i's are relatively small, e.g., the first row of Table 2, the lead time variances are relatively large and, consequently, the upper bounds $\tilde{s}_k$ are high. In these cases, the search-based solutions may be inferior to the DP solutions because

Table 2 Comparison Between the Solutions Found by the DP and Search Algorithm

		$\{s_1, s_2, s_5, s_6, s_9\}$				
$E(P_9)$	$\{n^1, n^2, n^5, n^6, n^9\}$	The DP solution	The search solution	Cost (DP)	Cost (search)	Percentage difference in costs (%)
7	{1, 2, 3, 2, 1}	{0, 0, 1, 0, 25}	{0, 0, 2, 0, 25}	41.89 ± 0.39	42.71 ± 0.4	−1.92
7	{2, 4, 5, 4, 2}	{0, 0, 1, 0, 21}	{1, 0, 4, 3, 18}	28.65 ± 0.32	29.14 ± 0.32	−1.68
7	{4, 7, 9, 8, 4}	{0, 0, 5, 3, 15}	{1, 1, 5, 4, 14}	20.99 ± 0.26	20.625 ± 0.26	1.77
7	{6, 10, 13, 11, 6}	{0, 0, 5, 3, 14}	{2, 1, 1, 1, 16}	17.58 ± 0.24	17.05 ± 0.23	3.11
1	{1, 2, 3, 2, 1}	{0, 0, 9, 6, 7}	{5, 4, 2, 5, 9}	27.4 ± 0.24	27.50 ± 0.24	−0.36
1	{2, 4, 5, 4, 2}	{3, 2, 5, 4, 6}	{4, 3, 4, 6, 5}	18.3 ± 0.19	17.87 ± 0.19	2.4
1	{4, 7, 9, 8, 4}	{0, 0, 7, 5, 4}	{3, 3, 3, 4, 5}	12.39 ± 0.15	12.28 ± 0.15	0.9
1	{6, 10, 13, 11, 6}	{0, 0, 6, 5, 4}	{0, 0, 5, 4, 5}	10.67 ± 0.14	10.51 ± 0.14	1.6

the search-based algorithm only evaluates discrete points in the base-stock level vector space. However, when the n^i's are large, e.g., the last row of Table 2, the lead-time variances are relatively small and, therefore, the upper bounds $\tilde{s}_k$ are low. As a result, the search-based algorithm can search the base-stock level vector space more exhaustively. In addition, because the DP algorithm is based on approximations, it may not be sensitive enough to small performance differences between various solutions. Thus, in these cases, the search-based algorithm can yield higher quality solutions than the DP algorithm.

6.2. Impact of Stock Positioning

Theorem 3.2 and Proposition 3.3 imply that increasing stock levels upstream while maintaining the overall stock level in a serial supply chain, results in longer backorder delays to external customers and a longer waiting time for any item in the system. On the other hand, using a computational study, Gallego and Zipkin (1999) illustrated that the impact of stock positioning is relatively small given that the overall stock level is about right. Motivated by these seemingly contradictory statements, in this section, we study the following issues: What is the magnitude of the impact of stock positioning on system performance? And, under what conditions is system performance sensitive to stock positions?

To answer these questions, we consider a simple two-stage serial system with stage 1 facing external demand. Without loss of generality, we assume zero transportation lead times and zero external replenishment lead times, but nonzero and constant processing times at each stage. By Definition 3.1 and Equations (3), (5), and (6), the backorder delay at stage 1 can be expressed by

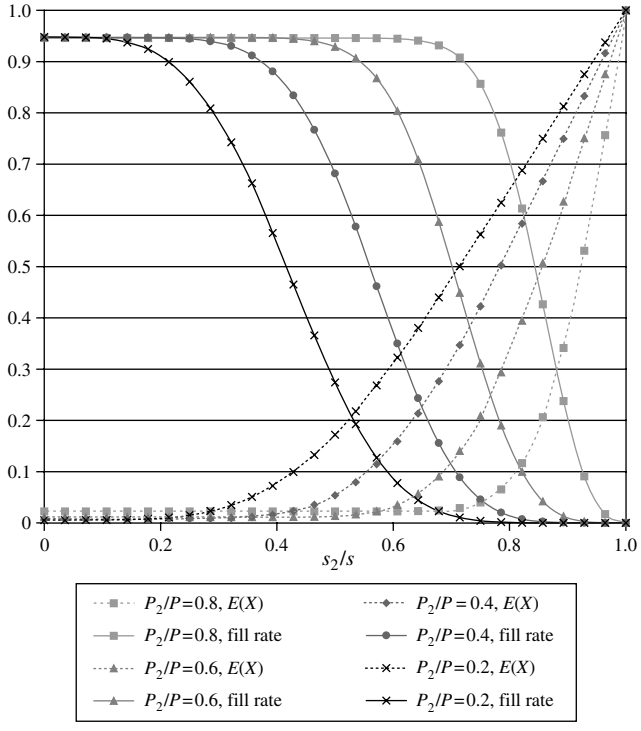
$$\begin{aligned} X_1 &= ((P_2 - T(s_2))^+ + P_1 - T(s_1))^+ \\ &= (P_2 + P_1 - T(s_2) - T(s_1) + (T(s_2) - P_2)^+)^+ \\ &= (P_1 + P_2 - T(s_1 + s_2) + (T(s_2) - P_2)^+)^+. \end{aligned} \quad (42)$$

First, note that $(T(s_2) - P_2)^+ = W_2$ is the inventory holding time at stage 2. Second, if we set $s_2 = 0$, then $X_1 = (P_1 + P_2 - T(s_1 + s_2))^+$. Therefore the extra backorder delay at $s_2 \neq 0$ relative to $s_2 = 0$ comes directly from the extra inventory holding time at stage 2. Intuitively, part of the additional inventory holding time at the upstream stage, W_2, is translated into the additional backorder delay at the downstream stage.

For simplicity, let $s = s_1 + s_2$ and $P = P_1 + P_2$. Figure 5 demonstrates the impact of P_2/P and s_2/s on the fill rate and expected backorder delay to external customers. In this example, $P = 20$, $\lambda = 1$, $\tau_1 = 0$, and $\alpha = 0.95$. The solid lines represent the fill rates as functions of s_2/s for different P_2/P, while the dashed lines represent the $E(X_1)$'s (normalized by the expected backorder delay at $s_2 = s$) for different P_2/P. Monte-Carlo simulation with sample size 40,000 is used to generate the results. Figure 1 reveals that

- When P_2/P is relatively high (e.g., 0.8), the fill rate and $E(X_1)$ are insensitive to s_2/s for a relatively wide range of s_2/s, e.g., when s_2/s varies from 0 to 0.7, $E(X_1)$ does not change much; similarly when s_2/s varies from 0 to 0.6, the fill rate remains almost the same.
- As P_2/P decreases, system performance becomes more sensitive to stock positioning. For instance, at $P_2/P = 0.2$, $E(X_1)$ does not change much only in the range $(0, 0.3)$, while the file rate does not change much only in $(0, 0.1)$.

Figure 5 The Impact of Stock Positions on System Performances



Intuitively, if P_2/P is relatively high (e.g., close to one), one has large flexibility in positioning the stocks as long as the overall stock level is correct. However, if P_2/P is relatively small (e.g., close to zero), then system performances, i.e., the fill rates and expected delays, become very sensitive to stock positions.

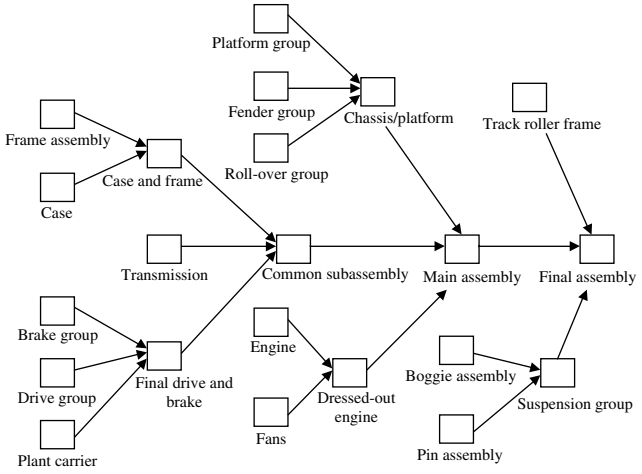
6.3. Impact of Lead-Time Uncertainty

In this section, we consider a more elaborate example; namely, Example 2, with 22 nodes and 21 arcs (see Figure 6). This example is inspired by a real-world problem, the Bulldozer supply chain (Graves and Willems 2003b). The objective of this section is to develop insights into the impact of the lead-time uncertainties on the system performances.

We refer the reader to Table 5 in Appendix C for the cost and lead-time data. While we keep the inventory holding costs and the expected processing times the same as in Graves and Willems (2003b), we assume stochastic processing times and stochastic, nonzero transportation lead times with the means generated randomly according to $Uniform\{1, 2, \ldots, 10\}$. The external supply lead times are zero. Without loss of generality, we assume that the external demand follows the Poisson process with $\lambda = 1$, and the target customer service at the final assembly is specified by $\tau = 0$ and $\alpha = 95\%$.

To study the impact of the lead-time uncertainties, we assume that all processing times and transportation lead times follow Erlang distributions with possibly different means but the same coefficient of variation, i.e., the same n. While this example is computationally prohibitive for the simulation-based search algorithm (§6.1), it takes around 20 seconds for the DP algorithm to generate a solution on a Pentium 1.67 GHz laptop.

Figure 6 The Example 2



First, we demonstrate the accuracy of the approximations in this example. For different values of n, we use the DP algorithm to determine the base-stock level at each node and then use simulation to estimate the total cost and fill rate. Table 3 presents the absolute percentage difference in costs between the simulation and the approximation, and the absolute difference in fill rates between simulation and the target. The confidence intervals for all simulated fill rates are no larger than 1.4%. The table shows that the cost approximations are very accurate for all lead-time c.v.s. On the other hand, the fill-rate approximations are reasonably accurate when the lead-time c.v.s. are relatively small. However, when the lead-time c.v.s. are relatively large, e.g., larger than 0.33, the fill-rate approximation may perform quite poorly. This is explained as follows: as we discussed in §6.1, normal distribution is a poor fit for the random lead times when n is close to 1. Indeed, when $n = 1$, the lead-time distribution is exponential, and the normal approximation can be far from accurate.

Second, we study the impact of lead-time uncertainties. Table 4 provides information on cost, overall stock levels, and fill rates as a function of the lead-time uncertainty. For each value of the lead-time c.v., we use the DP to determine the base-stock level at each stage. We then apply simulation to verify the service level provided to the external customers. If the service level did not closely match with the target, we adjusted the DP input fill rate until the simulated fill rate matched the 95% target. Columns "Cost (by simulation)" and "Sum of stock levels" provide information on total cost and the overall stock level in each case. We explain the remaining columns later. The last

Table 3 The Accuracy of the Approximations in Example 2

n	2	4	9	16	64	N/A
Lead-time c.v.	0.707	0.5	0.33	0.25	0.125	0
Absolute percentage difference in costs (%)	2.09	1.19	0.31	0.97	0.99	1.17
Difference in fill rates (%)	8.37	5.56	3.87	2.11	0.82	0.46

Table 4 The Impact of Lead-Time Uncertainty

n	Lead-time c.v.	Cost (by simulation)	Sum of stock levels	Simulated fill rate of the solution of constant lead times	Percentage cost increase (%)
2	0.707	$\$906,135 \pm 5,381$	375	0.285 ± 0.0088	25.51
4	0.5	$\$666,386 \pm 4,417$	320	0.456 ± 0.0098	11.64
9	0.33	$\$496,430 \pm 3,775$	241	0.664 ± 0.0093	7.8
16	0.25	$\$427,056 \pm 3,504$	222	0.785 ± 0.0081	6.15
64	0.125	$\$352,778 \pm 3,260$	191	0.902 ± 0.0058	0.74
N/A	0	$\$309,352 \pm 3,082$	185	0.949 ± 0.0043	N/A

row in Table 4 represents a system with deterministic processing times and transportation lead times.

Clearly, lead-time uncertainties have substantial impact on both systemwide inventory cost and stock levels, e.g., the cost is nearly tripled and the sum of stock levels is more than doubled as the coefficients of variation of all lead times increase from 0 to 0.707. The proposed approach can adjust the stock levels at each node according to the lead-time uncertainties. Table 6 in Appendix C lists the solutions found by the DP algorithm for different n. The table reveals that the lead-time uncertainties may also have an impact on the stock positions. While the stock levels at many nodes increase more or less proportionally as the lead-time uncertainties increase, the stock levels at nodes "Main assembly," "Common subassembly," "Chassis/platform," and "Dressed-out engine" may change dramatically as the uncertainties increase up to a certain level.

Our computational study also illustrates that ignoring lead-time uncertainties can lead to substantial errors. To provide measures of the potential errors, we replace the stochastic lead times with their means and use the DP algorithm to find a solution (note that we still allow for stochastic backorder delays). We then apply this solution to the original problem and determine the actual fill rate using simulation. This is listed in the fifth column "Simulated fill rate" of Table 4. Clearly, the policy based on a deterministic approximation of lead time may yield a customer service level far below the target level, especially when the lead-time uncertainties are quite significant.

Finally, if we set the DP input fill rates high enough so that the simulated fill rates of the solutions that ignore the lead-time uncertainties match with the target (95%), then the systemwide costs could be much higher than those of the solutions that incorporate the lead-time uncertainties. The last column of Table 4 lists the *percentage cost increase* of the solutions that ignore the uncertainties relative to the solutions that incorporate the uncertainties. We observe that the percentage cost increase can be significant for situations with relatively large lead-time uncertainties.

7. Conclusion

In this paper, we provide a unified framework to evaluate and coordinate inventory policies for supply chains with tree structures, where the lead times are stochastic, sequential, and exogenously determined. Each stage controls inventory with a continuous review base-stock policy, and the external demands follow independent Poisson processes. We follow the stochastic-service model approach, and derive recursive equations for the backorder delays at all stages. The recursive equations allow us to characterize the dependencies among different stages in the supply chain and to develop analytical and numerical insights with respect to safety stock positioning. A numerical study shows that the approximations are reasonably accurate for a wide range of parameters, and the algorithm based on dynamic programming can find the optimal or close to optimal solutions efficiently.

It is appropriate to conclude this paper by identifying the limitations of our framework, which also indicate future research directions.

- In this paper, we assume that external demand follows independent Poisson processes. Indeed, the recursive equations are derived for point processes, among which the compound Poisson processes are special cases. Thus Theorem 3.2 and Propositions 3.3, 3.4, and 3.8 hold for systems facing compound Poisson demand processes. However, compound Poisson

processes present additional challenges for performance evaluation and optimization because Proposition 3.9 may not hold, and the probability distribution of the backorder delay at each node may depend on the demand index.

- Our framework ignores production capacity limitations and batch size constraints. The evaluation and optimization of supply chains that include these constraints remains a challenging area.
- Extending this framework to supply chains with a cyclic networks is an important challenge.
- If the lead-time correlations in assembly systems cannot be ignored (as in ATO systems), then it is difficult to implement the DP algorithm, because incorporating the lead-time correlations will result in a significant increase in the dimensions of the state space. However, global optimization algorithms with decision variables either s_k or $E(X_k)$ can be implemented, and the DP algorithm that ignores the lead-time correlations can be used to generate a good starting point.
- Finally, better and more robust approximations are needed for supply chains with highly uncertain lead times.

Acknowledgments
The authors would like to thank the Senior Editor and two anonymous referees for their constructive comments and valuable suggestions, which have allowed them to improve the analytical and numerical results. Research supported in part by the Singapore-MIT Alliance; ONR contracts N00014-95-1-0232 and N00014-01-1-0146; and by National Science Foundation Contracts DMI-9732795, DMI-0085683, and DMI-0245352.

Appendix A

To prove Theorem 3.2, we first need the following technical Lemma 7.1.

Lemma 7.1. *Given any constants a, b, a', b', and c, d, if $a+b=a'+b'$ and $a\le a'$, then*

$$\xi = a - c + (b-d)^+ \le \xi' = a' - c + (b'-d)^+, \tag{43}$$

$$\eta = c - a + (d-b)^+ \ge \eta' = c - a' + (d-b')^+. \tag{44}$$

Proof. The case of $a=a'$ is trivial, so we focus on the case of $a<a'$. Clearly, $a<a'$ and $a+b=a'+b'$ imply $b>b'$. We consider the following three cases:

Case 1: $b'\ge d$. In this case, $\xi = a-c+b-d$ and $\xi' = a'-c+b'-d$, thus $\xi=\xi'$; $\eta = c-a$, and $\eta' = c-a'$, thus $\eta>\eta'$.

Case 2: $b'<d<b$. In this case, $\xi = a-c+b-d$ and $\xi' = a'-c$, thus $\xi<\xi'$ because $a+b-a'-d=b'-d<0$; $\eta = c-a$ and $\eta' = c-a'+d-b'$, thus $\eta>\eta'$ because $a'+b'-a-d = b-d>0$.

Case 3: $d\ge b$. In this case, $\xi = a-c$ and $\xi' = a'-c$, thus $\xi<\xi'$; $\eta = c-a+d-b$ and $\eta' = c-a'+d-b'$, thus $\eta=\eta'$. □

Proof of Theorem 3.2. It follows from Equations (3), (5), and (6) that

$$X_1(n_1) = [(L_2(n_2(n_1)-s_2) - T(s_2, n_2(n_1)))^+ + t_{2,1} + P_1 - T(s_1, n_1)]^+, \tag{45}$$

$$X_1'(n_1) = [(L_2(n_2'(n_1)-s_2') - T(s_2', n_2'(n_1)))^+ + t_{2,1} + P_1 - T(s_1', n_1)]^+. \tag{46}$$

Because systems with both stock positions are defined in the same probability space, it follows from the fact

$$n_2(n_1) - s_2 = n_1 - s_1 - s_2 = n_1 - s_1' - s_2' = n_2'(n_1) - s_2', \tag{47}$$

that in any event,

$$L_2(n_2(n_1)-s_2) = L_2(n_2'(n_1)-s_2'). \tag{48}$$

We also observe that in any event,

$$T(s_2, n_2(n_1)) + T(s_1, n_1) = T(s_2', n_2'(n_1)) + T(s_1', n_1), \tag{49}$$

$$T(s_1', n_1) \ge T(s_1, n_1). \tag{50}$$

Equation (49) follows from the fact that the expressions on both sides of this equation define the identical interarrival times (see Figure 1); inequality (50) is due to the fact that $s_1'>s_1$. Combining these observations with Lemma 7.1, it follows that in any event, $X_1'(n_1)\le X_1(n_1)$, i.e., $X_1'(n_1)\le_{\text{st}} X_1(n_1)$.

We now prove inequality (12). It follows from Equations (4)–(7) that

$$W_2(n_2(n_1)) + W_1(n_1) = (T(s_2, n_2(n_1)) - L_2(n_2(n_1)-s_2))^+ + [T(s_1, n_1) - (L_2(n_2(n_1)-s_2) - T(s_2, n_2(n_1)))^+ - t_{2,1} - P_1]^+, \tag{51}$$

$$W_2'(n_2'(n_1)) + W_1'(n_1) = (T(s_2', n_2'(n_1)) - L_2(n_2'(n_1)-s_2'))^+ + [T(s_1', n_1) - (L_2(n_2'(n_1)-s_2') - T(s_2', n_2'(n_1)))^+ - t_{2,1} - P_1]^+. \tag{52}$$

We analyze the following three cases:

Case 1: $T(s_2', n_2'(n_1)) \ge L_2(n_2'(n_1)-s_2')$. In this case, Equations (48)–(49) and inequality (50) imply that $T(s_2, n_2(n_1)) \ge L_2(n_2(n_1)-s_2)$. Hence, Equations (51)–(52) can be reduced to

$$W_2(n_2(n_1)) + W_1(n_1) = T(s_2, n_2(n_1)) - L_2(n_2(n_1)-s_2) + [T(s_1, n_1) - t_{2,1} - P_1]^+,$$

$$W_2'(n_2'(n_1)) + W_1'(n_1) = T(s_2', n_2'(n_1)) - L_2(n_2'(n_1)-s_2') + [T(s_1', n_1) - t_{2,1} - P_1]^+.$$

It follows from Equations (48)–(49), inequality (50), and Lemma 7.1 that $W_2'(n_2'(n_1)) + W_1'(n_1) \le W_2(n_2(n_1)) + W_1(n_1)$.

Case 2: $T(s_2', n_2'(n_1)) < L_2(n_2'(n_1) - s_2') < T(s_2, n_2(n_1))$. In this case, Equation (48) implies that Equations (51)–(52) can be reduced to

$$\begin{aligned} W_2(n_2(n_1)) + W_1(n_1) &= (T(s_2, n_2(n_1)) - L_2(n_2(n_1) - s_2))^+ \\ &\quad + [T(s_1, n_1) - t_{2,1} - P_1]^+, \\ W_2'(n_2'(n_1)) + W_1'(n_1) &= [T(s_2', n_2'(n_1)) - L_2(n_2'(n_1) - s_2') \\ &\quad + T(s_1', n_1) - t_{2,1} - P_1]^+. \end{aligned}$$

It follows from Equations (48)–(49) that $W_2'(n_2'(n_1)) + W_1'(n_1) \le W_2(n_2(n_1)) + W_1(n_1)$.

Case 3: $T(s_2, n_2(n_1)) \le L_2(n_2'(n_1) - s_2')$. In this case, Equation (49) and inequality (50) imply that $T(s_2', n_2'(n_1)) \le L_2(n_2'(n_1) - s_2')$. It follows from Equation (48) that Equations (51)–(52) can be reduced to

$$\begin{aligned} W_2(n_2(n_1)) + W_1(n_1) &= [T(s_2, n_2(n_1)) - L_2(n_2(n_1) - s_2) \\ &\quad + T(s_1, n_1) - t_{2,1} - P_1]^+, \\ W_2'(n_2'(n_1)) + W_1'(n_1) &= [T(s_2', n_2'(n_1)) - L_2(n_2'(n_1) - s_2') \\ &\quad + T(s_1', n_1) - t_{2,1} - P_1]^+. \end{aligned}$$

It follows from Equations (48)–(49) that $W_2'(n_2'(n_1)) + W_1'(n_1) = W_2(n_2(n_1)) + W_1(n_1)$. □

Proof of Proposition 3.3. First, note that $W_k'(n_k'(n_1)) = 0$ for all $k > 1$; therefore the system with SP′ can be reduced to a single-stage system with the total replenishment lead time equal to $S_K + \sum_{k=1}^{K} P_k + \sum_{k=1}^{K-1} t_{k+1,k}$. The proof is based on the following algorithm that transforms SP to SP′.

Step 1. If $s_2 = 0$, go to Step 2; otherwise, reduce s_2 by one and increase s_1 by one until s_2 becomes zero. Because $\sum_{k=3}^{K} W_k(n_k(n_1))$ remain stochastically the same, it follows from Theorem 3.2 that $X_1(n_1)$ and $\sum_{k=1}^{K} W_k(n_k(n_1))$ either become stochastically smaller or remain stochastically the same in each of these operations.

Step 2. Combine node 2 with node 1, so that the new node 1 has a processing time $P_1 + P_2$ and a transportation lead-time $t_{2,1} + t_{3,2}$. Renumber other nodes upstream sequentially by $2, 3, \ldots$. If all nodes other than the node 1 have zero base-stock levels, stop; otherwise, go back to Step 1.

Clearly, the output of the algorithm is the system with the stock positioning SP′. □

Proof of Proposition 3.4. It follows from Equation (15) that

$$X_k(n_k) = \left(\max_{i=1,2,\ldots,I}\{X_i(n_i(n_k)) + t_{i,k}\} + P_k - T(s_k, n_k)\right)^+, \quad (53)$$

$$X_k'(n_k) = \left(\max_{i=1,2,\ldots,I}\{X_i'(n_i'(n_k)) + t_{i,k}\} + P_k - T(s_k', n_k)\right)^+, \quad (54)$$

where

$$X_i(n_i(n_k)) = (L_i(n_i(n_k) - s_i) - T(s_i, n_i(n_k)))^+, \quad (55)$$

$$X_i'(n_i'(n_k)) = (L_i(n_i'(n_k) - s_i') - T(s_i', n_i'(n_k)))^+. \quad (56)$$

Because of the identical demand processes at nodes $i = 1, 2, \ldots, I$ and node k, we drop the subscripts from the T's. Choosing the indices appropriately yields $n_i(n_k) - s_i = n_i'(n_k) - s_i'$, and therefore $L_i(n_i(n_k) - s_i) = L_i(n_i'(n_k) - s_i')$ for all i.

Without loss of generality, let $X_1'(n_1'(n_k)) + t_{1,k} = \max_{i=1,2,\ldots,I}\{X_i'(n_i'(n_k)) + t_{i,k}\}$. If $X_1(n_1(n_k)) + t_{1,k} = \max_{i=1,2,\ldots,I}\{X_i(n_i(n_k)) + t_{i,k}\}$, then the problem is reduced to that of the serial systems, and the result follows immediately from Theorem 3.2. Otherwise, we must have

$$\begin{aligned} X_k'(n_k) &= (X_1'(n_1'(n_k)) + t_{1,k} + P_k - T(s_k', n_k))^+ \\ &\le (X_1(n_1(n_k)) + t_{1,k} + P_k - T(s_k, n_k))^+ \\ &\le X_k(n_k), \end{aligned}$$

where the first inequality follows from Theorem 3.2, and the second inequality is due to Equation (53). □

Proof of Proposition 3.7. Starting at the arrival time of demand n_k, we count backward all the previous interarrival times of external demands. Because of the independent Poisson processes, the interarrival times are mutually independent, and therefore they are associated (Tong 1980, Theorem 5.2.2 Part (d)). Furthermore, $X_i(n_i(n_k))$ are nonincreasing functions in each of these interarrival times because $L_i(s_i, n_i(n_k))$ are nondecreasing functions of these interarrival times. Conditioning on all the transit times, i.e., the processing cycle times, transportation lead times, and external lead times, it follows from Tong (1980, Theorem 5.2.3) and Assumption 1.1 that $X_i(n_i(n_k))$ are associated random variables. By Tong (1980, Theorem 5.2.4), inequality (25) holds for each n_k and τ and for each realizations of the transit times. Unconditioning on the transit times yields the desired result. □

Proof of Proposition 3.8. It follows from Equations (26), (28), and (29), and the assumption of the common probability space that

$$\begin{aligned} X_j(n_j) &= [(X_i(n_i(n_j)) + t_{i,k} + P_k - T_k(s_k, n_k(n_j)))^+ \\ &\quad + t_{k,j} + P_j - T_j(s_j, n_j)]^+, \\ X_j'(n_j) &= [(X_i(n_i(n_j)) + t_{i,k}' + P_k - T_k(s_k, n_k(n_j)))^+ \\ &\quad + t_{k,j}' + P_j - T_j(s_j, n_j)]^+. \end{aligned}$$

Combining Equation (31) and inequality (32), it follows from Lemma 7.1 that in any event, $X_j'(n_j) \le X_j(n_j)$, i.e., $X_j'(n_j) \le_{\text{st}} X_j(n_j)$.

We now prove inequality (34). Equations (30), (27), and (28) imply that

$$W_k(n_k(n_j)) = (T_k(s_k, n_k(n_j)) - X_i(n_i(n_j)) - t_{i,k} - P_k)^+, \quad (57)$$

$$\begin{aligned} W_j(n_j) &= [T_j(s_j, n_j) - t_{k,j} - P_j - (X_i(n_i(n_j)) \\ &\quad + t_{i,k} + P_k - T_k(s_k, n_k(n_j)))^+]^+. \end{aligned} \quad (58)$$

$W_k'(n_k(n_j))$ and $W_j'(n_j)$ can be expressed by Equations (57)–(58) if one replaces $t_{i,k}$ and $t_{k,j}$ by $t_{i,k}'$ and $t_{k,j}'$,

respectively. We analyze the following three cases:

Case 1: $W_k'(n_k(n_j)) > 0$. In this case, Equations (31), (57), and inequality (32) imply that $W_k(n_k(n_j)) > 0$. Hence

$$\begin{aligned} W_k(n_k(n_j)) + W_j(n_j) &= T_k(s_k, n_k(n_j)) - X_i(n_i(n_j)) - t_{i,k} \\ &\quad - P_k + [T_j(s_j, n_j) - t_{k,j} - P_j]^+, \\ W_k'(n_k(n_j)) + W_j'(n_j) &= T_k(s_k, n_k(n_j)) - X_i(n_i(n_j)) - t_{i,k}' \\ &\quad - P_k + [T_j(s_j, n_j) - t_{k,j}' - P_j]^+. \end{aligned}$$

It follows from Equation (31), inequality (32), and Lemma 7.1 that $W_k'(n_k(n_j)) + W_j'(n_j) \le W_k(n_k(n_j)) + W_j(n_j)$.

Case 2: $W_k'(n_k(n_j)) = 0$ and $W_k(n_k(n_j)) > 0$. In this case, we have

$$\begin{aligned} W_k(n_k(n_j)) + W_j(n_j) &= [T_j(s_j, n_j) - t_{k,j} - P_j]^+ + [T_k(s_k, n_k(n_j)) \\ &\quad - X_i(n_i(n_j)) - t_{i,k} - P_k]^+, \\ W_k'(n_k(n_j)) + W_j'(n_j) &= [T_j(s_j, n_j) - t_{k,j}' - P_j + T_k(s_k, n_k(n_j)) \\ &\quad - X_i(n_i(n_j)) - t_{i,k}' - P_k]^+. \end{aligned}$$

It follows from Equation (31) that $W_k'(n_k(n_j)) + W_j'(n_j) \le W_k(n_k(n_j)) + W_j(n_j)$.

Case 3: $W_k(n_k(n_j)) = 0$. In this case, Equations (31), (57), and inequality (32) imply that $W_k'(n_k(n_j)) = 0$. Hence

$$\begin{aligned} W_k(n_k(n_j)) + W_j(n_j) &= [T_j(s_j, n_j) - t_{k,j} - P_j + T_k(s_k, n_k(n_j)) \\ &\quad - X_i(n_i(n_j)) - t_{i,k} - P_k]^+, \\ W_k'(n_k(n_j)) + W_j'(n_j) &= [T_j(s_j, n_j) - t_{k,j}' - P_j + T_k(s_k, n_k(n_j)) \\ &\quad - X_i(n_i(n_j)) - t_{i,k}' - P_k]^+. \end{aligned}$$

It follows from Equation (31) that $W_k'(n_k(n_j)) + W_j'(n_j) = W_k(n_k(n_j)) + W_j(n_j)$. □

Appendix B

In the algorithm of Graves and Willems (2000), all nodes are initially labeled. We summarize the following properties of the labeling system:

(1) Each node labeled in the first $K-1$ steps is adjacent to exactly one node with higher label. Let $P(k)$ be the node with higher label for node $k = 1, \ldots, K-1$. For node K, all adjacent nodes have lower label.

(2) Following Graves and Willems (2000), we define $\mathcal{N}_k$, $k = 1, \ldots, K$ to be the subset of nodes $\{1, 2, \ldots, k\}$ that are connected to k on the subgraph of $\{1, 2, \ldots, k\}$.

$$\mathcal{N}_k = \{k\} + \bigcup_{i<k,\,(i,k)\in\mathcal{A}} \mathcal{N}_i + \bigcup_{j<k,\,(k,j)\in\mathcal{A}} \mathcal{N}_j.$$

The dynamic programming algorithm follows the order of labels and solves the program **P** for the subgraph $\mathcal{N}_k$ at each node k for certain inputs from the node $P(k)$. In particular,

• If $P(k)$ is a downstream node, we define $F_k(E(X_k), V(X_k))$ to be the minimal cost in the subgraph $\mathcal{N}_k$ for a given pair of $E(X_k)$ and $V(X_k)$.

• If $P(k)$ is an upstream node, we define $G_k(E(X_{P(k)}), V(X_{P(k)}))$ to be the minimal cost in the subgraph $\mathcal{N}_k$ for a given pair of $E(X_{P(k)})$ and $V(X_{P(k)})$.

Thus, given $\bar{X}_k$ (mean and variance), $E(X_k)$ and demand rate at node k, the total cost in the subgraph of $\mathcal{N}_k$ can be expressed by

$$\begin{aligned} C_k(E(\bar{X}_k), V(\bar{X}_k), E(X_k)) &= H_k(E(\bar{X}_k), V(\bar{X}_k), E(X_k)) \\ &\quad + \sum_{i<k,\,(i,k)\in\mathcal{A}} F_i(E(X_i), V(X_i)) \\ &\quad + \sum_{j<k,\,(k,j)\in\mathcal{A}} G_j(E(X_k), V(X_k)), \end{aligned} \tag{59}$$

where H_k and $V(X_k)$ are calculated by Equations (39)–(40), respectively. The first term is the average inventory carrying cost at node k, the second term represents the minimum costs for the suppliers of node k with lower labels, and the third term represents the minimum costs of the immediate customers of node k with lower labels.

For a node $k = 1, 2, \ldots, K-1$, if $P(k)$ is the downstream of k, then for each feasible pair of $E(X_k)$ and $V(X_k)$ that satisfies $E(X_k) \in [0, Q_k]$ and $V(X_k) \in [0, \sigma_k^2]$,

$$\begin{aligned} &F_k(E(X_k), V(X_k)) \\ &\quad = \min_{E(X_i),\,V(X_i),\,(i,k)\in\mathcal{A}} \{C_k(E(\bar{X}_k), V(\bar{X}_k), E(X_k))\} \\ &\qquad \text{s.t.} \quad L_k = \max\{X_i + t_{i,k}, \forall (i,k) \in \mathcal{A}; S_k\} + P_k, \\ &\qquad\qquad E(X_k) \le E(L_k), \\ &\qquad\qquad V(X_k) = f_k(E(\bar{X}_k), V(\bar{X}_k), E(X_k)). \end{aligned} \tag{60}$$

Clearly, each pair of $E(X_i)$ and $V(X_i)$ can only be chosen from its feasible region.

If $P(k)$ is the upstream of k, then for each feasible pair of $E(X_{P(k)})$ and $V(X_{P(k)})$ that satisfies either $E(X_{P(k)}) = 0$ and $V(X_{P(k)}) = 0$ or $0 < E(X_{P(k)}) \le Q_{P(k)}$ and $0 < V(X_{P(k)}) \le \sigma_{P(k)}^2$,

$$\begin{aligned} &G_k(E(X_{P(k)}), V(X_{P(k)})) \\ &\quad = \min_{E(X_i),\,V(X_i),\,i<k,\,(i,k)\in\mathcal{A};\,E(X_k)} \{C_k(E(\bar{X}_k), V(\bar{X}_k), E(X_k))\} \\ &\qquad \text{s.t.} \quad L_k = \max\{X_i + t_{i,k}, \forall (i,k) \in \mathcal{A}; S_k\} + P_k, \\ &\qquad\qquad E(X_k) \le \min\{Q_k, E(T_k)\}, \\ &\qquad\qquad V(X_k) = f_k(E(\bar{X}_k), V(\bar{X}_k), E(X_k)) \le \sigma_k^2. \end{aligned} \tag{61}$$

Each pair of $E(X_i)$ and $V(X_i)$, $i < k$, $(i,k) \in \mathcal{A}$ choose a value from its feasible region.

Now, we are ready to describe the dynamic program:

(1) For $k = 1$ to $K-1$, if $P(k)$ is the downstream node of k, then we evaluate $F_k(E(X_k), V(X_k))$ by Equation (60) for each feasible discrete pair of $E(X_k)$ and $V(X_k)$. If $P(k)$ is the upstream node of k, then we evaluate $G_k(E(X_{P(k)}), V(X_{P(k)}))$ by Equation (61) for each feasible discrete pair of $E(X_{P(k)})$ and $V(X_{P(k)})$.

(2) For $k = K$, we evaluate $F_K(E(X_K), V(X_k))$ for each feasible discrete pair of $E(X_K)$ and $V(X_K)$, and then we pick the minimum.

All optimizations are solved by enumeration. To identify the computational complexity of the algorithm, we define $\delta_k = \sigma_k^2$ if $P(k)$ is the downstream of node k, and $\delta_k = 1$ otherwise. For node k, the computation effort is at most proportional to $O(Q_k \times \delta_k \times \prod_{(i,k)\in\mathcal{M}}(Q_i \times \sigma_i^2))$, and the total computational effort is at most proportional to $O(\sum_{k=1}^{K} Q_k \times \delta_k \times \prod_{(i,k)\in\mathcal{M}}(Q_i \times \sigma_i^2))$. Thus the major computational complexity comes from the assembly nodes.

To illustrate the DP algorithm, we consider a three-stage serial system with the following labels: node 1 supplies node 3, which supplies node 2. This system can be solved as follows:

(1) For node 1, we first note that $V(X_1)$ is uniquely determined by $E(X_1)$. Therefore, for each feasible $E(X_1) \in [0, Q_1]$, we identify the corresponding base-stock level s_1 and calculate $F_1(E(X_1), V(X_1))$, where $V(X_1)$ depends on $E(X_1)$.

(2) For node 2 and each pair of $E(X_3) \in [0, Q_3]$ and $V(X_3) \in [0, \sigma_3^2]$, we determine $G_2(E(X_3), V(X_3))$ by identifying the optimal $E(X_2)$ (or s_2, equivalently), so that the inventory cost of node 2 is minimized and all constraints in Equation (61), and the service constraint are satisfied.

(3) For node 3 and each feasible pair of $E(X_3) \in [0, Q_3]$ and $V(X_3) \in [0, \sigma_3^2]$, we compute $F_3(E(X_3), V(X_3))$ by enumerating all feasible $E(X_1)$. Then, we identify the pair of $E(X_3)$ and $V(X_3)$ with minimal $F_3(E(X_3), V(X_3))$.

Appendix C

Table 5 Data for Example 2

Node i	$E(P_i)$	Added cost ($)	Node j	$E(t_{i,j})$
Platform group	6	725	Chassis/platform	7
Fender group	9	900	Chassis/platform	5
Roll-over group	8	1,150	Chassis/platform	4
Chassis/platform	7	4,320	Main assembly	7
Frame assembly	19	605	Case and frame	9
Case	15	2,200	Case and frame	4
Case and frame	16	1,500	Common subassembly	10
Brake group	8	3,850	Final drive and brake	6
Drive group	9	1,550	Final drive and brake	5
Plant carrier	9	155	Final drive and brake	2
Final drive and brake	6	3,680	Common subassembly	9
Engine	7	4,500	Dressed-out engine	10
Fans	12	650	Dressed-out engine	5
Dressed-out engine	10	4,100	Main assembly	2
Boggie assembly	11	575	Suspension group	1
Pin assembly	35	90	Suspension group	6
Suspension group	7	3,600	Final assembly	7
Transmission	15	7,450	Common subassembly	4
Common subassembly	5	8,000	Main assembly	5
Main assembly	8	12,000	Final assembly	7
Track roller frame	10	3,000	Final assembly	10
Final assembly	4	8,000	N/A	N/A

Table 6 Solutions for Example 2

Node i	$n = 2$	$n = 4$	$n = 9$	$n = 16$	Constant lead times
Platform group	4	4	2	0	0
Fender group	5	6	2	2	0
Roll-over group	3	2	0	0	0
Chassis/platform	19	18	1	1	0
Frame assembly	23	18	18	17	12
Case	2	0	0	0	0
Case and frame	38	35	32	31	30
Brake group	0	0	0	0	0
Drive group	7	5	3	2	0
Plant carrier	11	8	5	3	0
Final drive and brake	9	8	9	10	9
Engine	0	0	0	0	0
Fans	11	8	7	5	0
Dressed-out engine	20	21	1	2	1
Boggie assembly	3	4	5	5	3
Pin assembly	86	65	52	47	39
Suspension group	21	20	17	16	16
Transmission	0	0	0	0	0
Common subassembly	21	21	2	2	3
Main assembly	24	24	39	34	35
Track roller frame	18	15	13	12	11
Final assembly	50	38	33	33	26

We summarize implementation details as follows. From the above example and our numerical study, we observe that most pairs of $E(X_k)$ and $V(X_k)$ are infeasible in solving Equation (60). In addition, $E(X_k)$ is often more important than $V(X_k)$ in determining the system performance. Therefore we can significantly improve the numerical efficiency of the DP algorithm by substituting $F_k(E(X_k))$ for $F_k(E(X_k), V(X_k))$. More specifically, we compute F_k only over the feasible region of $E(X_k)$, and determine $V(X_k) \le \sigma_k^2$ for each $E(X_k)$ by $V(X_k) = f_k(E(\overline{X}_k^*), V(\overline{X}_k^*), E(X_k))$, where $\overline{X}_k^*$ are the backorder delays at the suppliers of node k that solve for $F_k(E(X_k))$ in Equation (60).

We discretize the time line so that the expected backorder delay and delay variance at each node are restricted to take only discrete values. Consider an arbitrary node k and assume that $P(k)$ is the downstream node. To determine the corresponding base-stock level at node k for a given $E(X_k)$, we choose the smallest integer so that the actual expected backorder delay at node k is smaller than or equal to $E(X_k)$. If the actual expected backorder delay at the base-stock level chosen does not match the given $E(X_k)$, we set the $E(X_k)$ to be the actual expected backorder delay. If $P(k)$ is the upstream node, then we need to enumerate all possible pairs of $E(X_{P(k)})$ and $V(X_{P(k)})$ (see Equation (61)). To improve numerical efficiency, the nodes in the networks are labeled in such a way that the number of nodes with upstream $P(k)$ is minimized. When we evaluate Equation (59), $E(X_k)$ and $V(X_k)$ may not completely match with the discrete values of $E(X)$ and $V(X)$ for which

G_j, $j < K$, $(k, j) \in \mathcal{A}$ is calculated. In this event, we either identify the closest discrete values of $E(X)$ and $V(X)$ or use interpolation.

References

Axsäter, S. 1993. Continuous review policies for multi-level inventory systems with stochastic demand. S. Graves, A. Rinnooy Kan, P. Zipkin, eds. *Logistics of Production and Inventory*. Elsevier Science Press, North-Holland, Amsterdam, The Netherlands.

Axsäter, S., K. Rosling. 1993. Notes: Installation vs. echelon stock policies for multilevel inventory control. *Management Sci.* **39** 1274–1280.

Boyaci, T., G. Gallego. 2001. Serial production/distribution systems under service constraints. *Manufacturing Service Oper. Management* **3** 43–50.

Caglar, D., C. L. Li, D. Simchi-Levi. 2004. Two-echelon spare parts inventory system subject to a service constraint. *IIE Trans.* **36** 655–666.

Clark, A. J., H. E. Scarf. 1960. Optimal policies for a multi-echelon inventory problem. *Management Sci.* **6** 475–490.

Clark, C. E. 1961. The greatest of a finite set of random variables. *Oper. Res.* **9** 145–162.

Diks, E. B., A. G. de Kok, A. G. Lagodimos. 1996. Multi-echelon systems: A service measure perspective. *Eur. J. Oper. Res.* **95** 241–263.

Esary, J. D., F. Proschan, D. W. Walkup. 1967. Association of random variables, with applications. *Ann. Math. Statist.* **38** 1466–1474.

Ettl, M., G. E. Feigin, G. Y. Lin, D. D. Yao. 2000. A supply network model with base-stock control and service requirements. *Oper. Res.* **48** 216–232.

Federgruen, A. 1993. Centralized planning models for multi-echelon inventory systems under uncertainty. S. C. Graves et al., eds. *Handbooks in Operations Research and Management Science*, Vol. 4. North-Holland, Amsterdam, The Netherlands, 133–173.

Feigin, G. E. 1999. Inventory planning in large assembly supply chains. S. Tayur, R. Ganeshan, M. J. Magazine, eds. *Quantitative Models for Supply Chain Management*. Kluwer Academic Publishers, Boston, MA.

Gallego, G., P. Zipkin. 1999. Stock positioning and performance estimation in serial production-transportation systems. *Manufacturing Service Oper. Management* **1** 77–88.

Glasserman, P., S. Tayur. 1995. Sensitivity analysis for base-stock levels in multiechelon production-inventory systems. *Management Sci.* **41** 216–232.

Glasserman, P., S. Tayur. 1996. A simple approximation for a multistage capacitated production-inventory systems. *Naval Res. Logist.* **43** 41–58.

Graves, S. 1985. Multi-echelon inventory model for a repairable item with one-for-one replenishment. *Management Sci.* **31** 1247–1256.

Graves, S. 1988. Safety stocks in manufacturing systems. *J. Manufacturing Oper. Management* **1** 67–101.

Graves, S. C., S. P. Willems. 2000. Optimizing strategic safety stock placement in supply chains. *Manufacturing Service Oper. Management* **2** 68–83.

Graves, S. C., S. P. Willems. 2003a. Erratum: Optimizing strategic safety stock placement in supply chains. *Manufacturing Service Oper. Management* **5** 176–177.

Graves, S. C., S. P. Willems. 2003b. Supply chain design: Safety stock placement and supply chain configuration. A. G. de Kok, S. C. Graves, eds. *Handbooks in Operations Research and Management Science*, Vol. 11, *Supply Chain Management: Design, Coordination and Operation*. North-Holland Publishing Company, Amsterdam, The Netherlands.

Hanssmann, F. 1959. Optimal inventory location and control in production and distribution networks. *Oper. Res.* **7** 483–498.

Janakiraman, G., J. A. Muckstadt. 2004. A decomposition approach for a class of capacitated serial systems. Working paper, Stern School of Business, New York University, New York.

Kapuscinski, R., S. Tayur. 1999. Optimal policies and simulation-based optimization for capacitated production inventory systems. S. Tayur, R. Ganeshan, M. J. Magazine, eds. *Quantitative Models for Supply Chain Management*. Kluwer Academic Publishers, Boston, MA.

Kulkarni, V. G. 1995. *Modeling and Analysis of Stochastic Systems*. Chapman and Hall, New York.

Lee, H. L., C. Billington. 1993. Material management in decentralized supply chains. *Oper. Res.* **41** 835–847.

Lee, H. L., K. Moinzadeh. 1987a. Two-parameter approximations for multi-echelon repairable inventory models with batch ordering policy. *IIE Trans.* **19** 140–149.

Lee, H. L., K. Moinzadeh. 1987b. Operating characteristics of a two-echelon inventory system for repairable and consumable items under batch ordering and shipment policy. *Naval Res. Logist. Quart.* **34** 365–380.

Lee, Y., P. Zipkin. 1992. Tandem queues with planned inventories. *Oper. Res.* **40** 936–947.

Lee, Y., P. Zipkin. 1995. Processing networks with planned inventories: Sequential refinement systems. *Oper. Res.* **43** 1025–1036.

Lin, G., M. Ettl, S. Buckley, S. Bagchi, D. D. Yao, B. L. Naccarato, R. Allan, K. Kim, L. Koenig. 2000. Extended-enterprise supply-chain management at IBM Personal Systems Group and other divisions. *Interfaces* **30** 7–25.

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

Palm, C. 1938. Analysis of the Erlang traffic formula of busy-signal arrangements. *Ericsson Technics* **5** 39–58.

Parker, R. P., R. Kapuscinski. 2004. Optimal policies for a capacitated two-echelon inventory systems. *Oper. Res.* **52** 739–755.

Porteus, E. L. 2002. *Foundations of Stochastic Inventory Theory*. Stanford University Press, Stanford, CA.

Schwartz, L., B. Deuermeyer, R. Badinelli. 1985. Fill-rate optimization in a one warehouse N-identical retailer distribution system. *Management Sci.* **31** 488–498.

Shaked, M., J. G. Shanthikumar. 1994. *Stochastic Orders and Their Applications*. Academic Press, New York.

Sherbrooke, C. 1968. METRIC: A multi-echelon technique for recoverable item control. *Oper. Res.* **16** 122–141.

Sherbrooke, C. 1986. VARI-METRIC: Improved approximations for multi-indenture, multi-echelon availability models. *Oper. Res.* **34** 311–319.

Simpson, K. F. 1958. In-process inventories. *Oper. Res.* **6** 863–873.

Song, J. S., P. Zipkin. 2002. Supply chain operations: Assemble-to-order systems. A. G. de Kok, S. C. Graves, eds. *Handbooks in Operations Research and Management Science*, Vol. 11. *Supply Chain Management*. Elsevier, Amsterdam, The Netherlands.

Speck, C., J. van der Wal. 1991. The capacitated multi-echelon inventory system with serial structure: 1. The "Push-Ahead" effect. Memorandum COSOR 91-39, Eindhoven University of Technology, Eindhoven, The Netherlands.

Svoronos, A., P. Zipkin. 1988. Estimating the performance of multi-level inventory systems. *Oper. Res.* **36** 57–72.

Svoronos, A., P. Zipkin. 1991. Evaluation of one-for-one replenishment policies for multiechelon inventory systems. *Management Sci.* **37** 68–83.

Tong, Y. L. 1980. *Probability Inequalities in Multivariate Distributions*. Academic Press, New York.

van Houtum, G. J., K. Inderfurth, W. H. M. Zijm. 1996. Materials coordination in stochastic multi-echelon systems. *Eur. J. Oper. Res.* **95** 1–23.

Zhao, Y., D. Simchi-Levi. 2005. Performance analysis and evaluation of assemble-to-order systems with stochastic sequential lead-times. *Oper. Res.* Forthcoming.

Zipkin, P. 1991. Evaluation of base-stock policies in multiechelon inventory systems with compound-Poisson demands. *Naval Res. Logist.* **38** 397–412.

Zipkin, P. 2000. *Foundations of Inventory Management*. McGraw-Hill, Boston, MA.