Interfaces
Vol. 36, No. 3, May–June 2006, pp. 259–271
ISSN 0092-2102 | EISSN 1526-551X | 06 | 3603 | 0259
DOI 10.1287/inte.1060.0212

Achieving Breakthrough Service Delivery Through Dynamic Asset Deployment Strategies

Morris A. Cohen
The Wharton School, University of Pennsylvania, 546 Jon M. Huntsman Hall, 3730 Walnut Street, Philadelphia, Pennsylvania 19104-6340, cohen@wharton.upenn.edu

Narendra Agrawal
Department of Operations and MIS, Leavey School of Business, Santa Clara University, Santa Clara, California 95053, nagrawal@scu.edu

Vipul Agrawal
MCA Solutions, 2 Penn Center Plaza, Suite 700, 1500 JFK Boulevard, Philadelphia, Pennsylvania 19102, vipul.agrawal@mcasolutions.com

Many firms have shifted their focus from their products to their customers and the value derived from owning and using the products. They see after-sales service as an important source of revenue and profit, customer acquisition and retention, and competitive differentiation. However, they also find it challenging to manage their service-supply chain. Service organizations must position and manage service-supply-chain resources optimally to support the delivery of after-sales service. They must also develop capabilities to respond rapidly to the demand for service in a cost-effective manner. To succeed in implementing a service-centric strategy, firms must determine what items in their products' service bill-of-material hierarchy should be deployed throughout their geographical hierarchy of service support locations. They must make these complex and interrelated decisions in anticipation of service demand, which is uncertain. Firms must also be flexible and should understand the mechanisms in a service-supply chain needed to fulfill customers' demands for service and the resulting demands for support assets and capacities. Dynamic asset deployment (DAD), a collection of management policies that promote this flexibility, can be used to develop the capabilities needed to effectively and profitably deliver services. These policies require a real-options-based optimization approach to decision making.

Key words: inventory, production: multi-item, multiechelon; industries: telecommunications.
History: This paper was refereed.

During the last decade, many firms have shifted their focus from their products to their customers and the value that derived from ownership and use of the products. In adopting this customer-centric perspective, firms are blurring the line between products and services. For example, Hewlett Packard and IBM historically competed as manufacturers of cutting-edge, high-technology products. Recently they have adopted a competitive strategy of providing technology solutions, in which services are a major component. Today, most firms would describe their output as a bundle of goods and services, and the proportion of their revenue derived from services is growing.

For example, in 2001, IBM derived over $5 billion of its revenue from maintenance and service. (IBM's 2005 sale of its PC manufacturing division to China's Lenovo group is further evidence of its shift towards a service-oriented strategy.) In a recent AMR report, Bijesse et al. (2002) put the percentage of revenue from services at 24 percent. An informal survey of participants at the Wharton-Stanford conference on the after-sales service industry in 2004 revealed that many traditional manufacturing firms derive over 50 percent of their revenues from services and expect this figure to continue to increase. Companies recognize that after-sales service is an important source of revenue and profit, customer acquisition and retention, and competitive differentiation. They are beginning to focus on their service-supply chains, the network of resources that include material (service parts), people (service engineers, call-center staff, repair staff, warehouse staff, and transportation staff), and infrastructure (for moving and storing materials, repair, transportation, information systems, and communication). Managing the service-supply chain is challenging, even though one might expect that lessons learned in managing the supply chains for manufactured goods would be directly applicable to service delivery. This is not so, because the mechanisms needed to design, produce and deliver postsales service differ significantly from those used for manufactured goods and require considerable investment in dedicated assets.

For example, Cisco Systems, a manufacturer of computer, data-storage, and communications-related equipment, has long recognized the importance of after-sales service to its overall growth strategy. With revenues of about $3.9 billion in 2005, Cisco has seen the revenues of its customer advocacy division (its service organization) grow by about 13 percent since 2004.

To provide customer support, Cisco offers a gamut of services, such as troubleshooting, hardware and software support, systems monitoring, and systems management. It does so through warranty services, requiring only parts, and through service contracts, requiring parts and service engineers. It must manage tens of thousands of service contracts that range from two-hour to 10-day delivery for return-to-factory services. Cisco collaborates with external partners for its repair, manufacturing, and logistics needs. Its service-supply chain consists of a worldwide network of service engineers, about 800 fulfillment depots, 18 repair centers, and five material-return-processing centers, which collectively deliver approximately 720,000 parts and repair approximately 530,000 parts annually.

Because of this complexity, service divisions of such companies as Cisco often have

—Less than desirable customer-service levels,

—An inability to respond to frequent changes in the installed base and customer-service contracts,

—High investments in slow moving parts inventory,

—Complex business processes,

—A need for manual oversight, and,

—A lack of regional and global coordination.

Consequently, service-support organizations must manage resources and business processes to fulfill various customer-support entitlements through their complex global network. They also must control operating costs and capital investments and meet rising customer expectations for service.

Such situations are common among companies in many industries, such as high technology, aerospace and defense, telecommunications, and automotive. The problem a service organization faces can be articulated in two questions:

(1) How should the organization optimally position and manage service-supply-chain resources to deliver after-sales service?

(2) What is the most cost-effective and service-effective way to provide service?

These questions are difficult to answer because service-supply chains are complex and face high demand and supply uncertainty. Therefore, a static approach to providing services, as provided by the enterprise resource planning (ERP) and distribution resource planning (DRP) systems used in manufacturing supply chains, is inefficient. Essential to a service-centric strategy is flexibility based on understanding the mechanisms for meeting customers' demands for service and maintaining the availability of support assets and capacities. We call the management policies that promote this flexibility dynamic asset deployment (DAD). In this hierarchical approach to asset management and deployment, the firm makes strategic and tactical decisions dynamically based on the most recent forecasts of supply and demand. At each step, it uses estimates of future uncertainties to position assets throughout the service-supply chain so that it can satisfy demands for service efficiently.

Service Asset Management

The after-sales service products customers buy are equivalent to entitlements for response to a support need within a specified time, at a specified level of reliability and for a specified price. A firm's service-supply chain must seek to maximize the benefit customers derive from owning and using the products they purchased. Firms cannot produce service products in advance and thus cannot store them on a shelf. They must produce them for immediate consumption, typically in response to a contingency, such as the failure of an installed product. Such service events are difficult to predict. To satisfy demand for service products, the organization relies on physical assets, such as spare-parts inventory, repair-depot capacity, and service engineers. It must deploy these assets in advance of service events to respond within the promised time (within minutes or hours).

Firms must deploy their service assets to maximize the profits they derive from providing diverse service products to a heterogeneous group of customers with geographically distributed installed equipment. They must deploy these assets to take advantage of the relationships among resources, decisions, and information that characterize service-supply chains. Companies must understand the trade-offs associated with delivering service, the strategic options for deploying various service-delivery assets, and the effect of asset-deployment decisions and fulfillment policies on service delivery trade-offs.

Trade-Offs for Service Delivery

The principal trade-off supply chain managers must consider is the trade-off between revenue, cost, and service performance.

A firm obtains revenue by selling service products designed to provide customers with uninterrupted use of its products. It can capture this revenue by selling performance-based service contracts or by charging for time (for technicians, labor, and for use of a repair depot) and material (for replacement parts). Many firms also meet their customers' expectation for service through product warranties that act as prepaid service contracts operating for a fixed period of time after the sale. Consequently, the revenue model for service should consider the initial sale price of the product, the terms of the warranty, the design and price of the contracts to cover service after the warranty expires, the prices for noncontracted time and material, and the impact of delivery service on repeat buying. Many firms reduce their initial product sale prices to capture the long-term revenue stream associated with delivering service (following the famous razor blade strategy in which firms distributed the product free of charge to capture the revenue stream of support products from captive users) (Gallagher et al. 2005). Others focus on designing their products and infrastructure to promote serviceability and to reduce the likelihood that customers will seek other sources of service in the after market (Cohen and Whang 1997 and Cohen et al. 2000). In all cases, the revenue service providers can expect is a function of the service products it delivers.

Service-delivery costs depend on how managers decide on the deployment and control of service resources. They must do so in a manner that meets specific service targets and satisfies budget and resource constraints. The direct costs of meeting demand for service resources include the costs of fault diagnosis, material handling, transportation, and parts repair. A variety of capital and fixed costs are also associated with providing service, for example, investment tied up in parts inventory and the fixed cost of warehouse and repair facilities. Finally, firms have costs for infrastructure, training, and human resources associated with providing service.

Service performance is directly related to product uptime. From the customer's perspective, service quality is defined by the delay between a request for service (for example, because of product failure) and their product's restoration to full operating condition at their facility. From a service-supply-chain provider's perspective, various measures are associated with the availability of the resources it needs to provide customer service. The principal internal metric is *part fill rate*, the fraction of demand for parts that is available in stock at the site receiving the demand. The primary measure used to capture the customer's perception of service quality is *product availability*, the fraction of time the product is available for use. Product availability is largely determined by the delay in meeting demand for resources throughout the service-supply chain. Performance under both classes of service metrics (fill-rate based and availability based) are functions of the underlying risk structure of future demand and the positioning and management of resources throughout the service-supply chain.

The trade-off between these metrics can be illustrated by an efficient-frontier curve. Because the demand for service is random, the timing, location, extent, and consequences of a demand for service cannot be forecast with certainty. Since the response times specified in service contracts typically are much shorter than the lead times for moving or acquiring material, and the timing of the demands is uncertain, firms must position many assets prior to service events. The response time required for servicing failed semiconductor equipment in wafer-fabrication facilities can be as fast as 15 minutes because the cost of downtime in fabs can be as high as $1 million per day. In general, the greater the level of service promised, the larger the required investment in such assets. Carrying little or no inventory of parts saves on financial costs. However, it means long response times for service events because parts must be ordered or repaired. While calculating the relationship between service performance and costs is nontrivial, it can be done with advanced analytical methods. We can depict the relationship with an efficient frontier curve (Figure 1). The costs increase disproportionately as the promised service level increases.

The service provider's problem is to determine a way to meet commitments for service performance (in terms of cost and speed) in the most effective manner as defined by the cost/service efficient frontier. To be inside the frontier suggests inefficiency. To be on the frontier suggests that increasing the level of service will also increase costs. Balancing the trade-offs among revenue, cost, and service is challenging because of escalating service expectations, service-supply-chain complexity, and the uncertainty associated with service events.

Service-Asset-Deployment Options

To explain how the revenue-cost-performance trade-off operates in the service-supply chain, we use a two dimensional hierarchy, based on supply chain geography (geographical hierarchy) and the product design architecture (product hierarchy). We use this geography-product hierarchy to describe how firms deploy and manage resources. In particular, for after-sales service, we will determine where to deploy in terms of the geographical hierarchy and what to deploy in terms of the product hierarchy.

The product hierarchy for any end product consists of major modules (field replaceable units (FRUs)), submodules (shop replaceable units (SRUs)), and parts (Figure 2). The information required to specify a product hierarchy can typically be found in the bill of material (BOM), the recipe for producing the product. The production BOM specifies the number of units of each input needed to produce every item in the product hierarchy. The service BOM, on the other hand, specifies the number of units of each input needed to repair every item in the product hierarchy (where replacement rates are generally fractional numbers because every repair instance of a parent part may not need the same quantity of a child part). Both production and service BOMs must incorporate changes introduced due to engineering changes and part supercessions (Table 1).

A geographical hierarchy organizes stocking locations (nodes) by echelon to capture material and information flows associated with fulfilling demand in the service-supply chain (Figure 3). The inventory

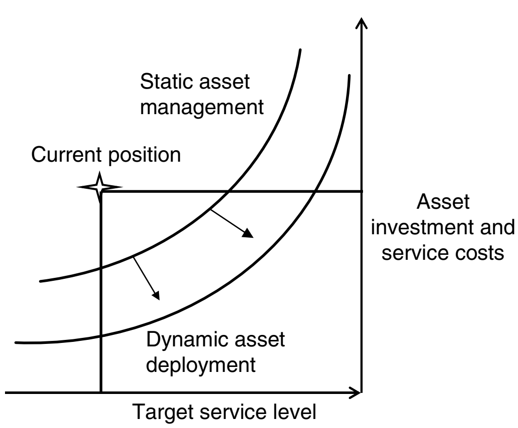

Figure 1: The efficient frontier curve shows how inventory investments and costs increase if the target service level increases. Dynamic asset management can help push the efficient frontier curve down.

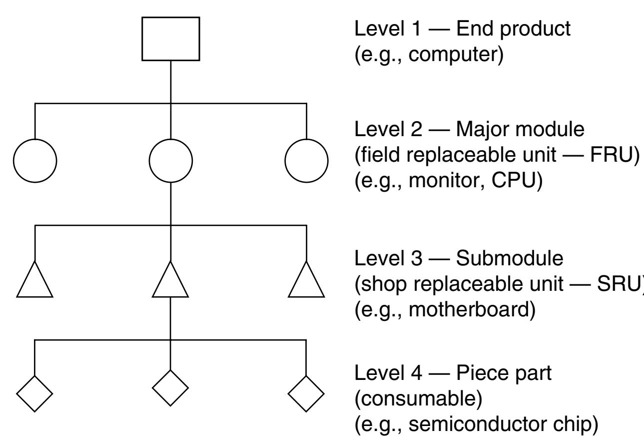

Figure 2: Each node represents an item in a product's hierarchical structure. Each layer in this hierarchy is assigned an indenture level, starting with one for end products and increasing by one as we traverse downwards.

Indenture	Part name: FRU	Part name: SRU	Part name: Piece part	Production BOM quantity	Part version	Service BOM quantity
1	Router				1	
2		Motherboard		1	1	0.1
3			REA-048	3	2	1.6
2		Power supply		1	1	0.1
3			145D0111-3	1	2	0
3			REA-048	3	1	2.1
2		Command unit		1	1	0
3			RLA27SM4-00	2	1	0
3			PLRA6560010	3	1	3
3			PLRA6560010	2	2	2
3			MS51989-105-10	1	1	0
3			145D0111-3	8	3	0
2		Interface unit		4	1	0
3			145D0121-3	1	1	0
2		Display unit		1	1	1
3			145DS011-10	6	1	2.2
3			145D0101-5	1	1	0

Table 1: In the repair of a router, the expected number of motherboards consumed is 0.1 (a replacement rate of 0.1). Similarly in the repair of the motherboard, part REA-048, which occurs three times in the assembly, is replaced at a rate of 1.6 units per repair. Parts whose replacement rate is zero have never been replaced (at this location) or by design are not replaced and a higher level component that contains that part is replaced.

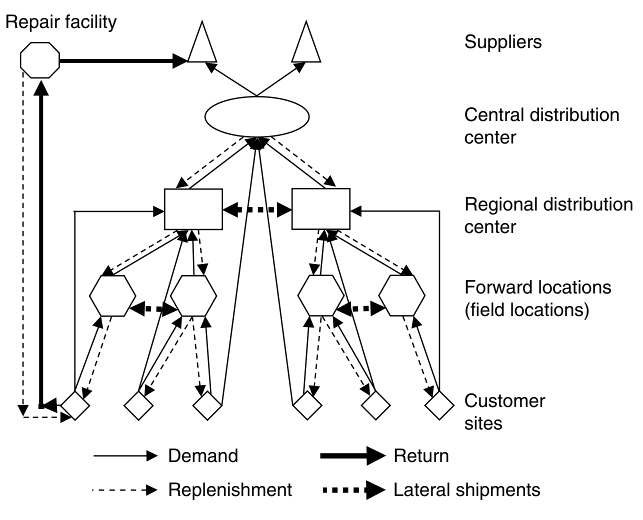

Figure 3: Sites at the lowest echelon, which could include customer sites, are called the forward or field locations. At the top of the hierarchy are the central distribution centers, the emergency backup and replenishment sites for the downstream locations. In between, the network could include additional regional distribution centers. Both emergency (customer) demand and replenishment demand flow up through the hierarchy and ultimately get replenished from suppliers. Reverse material flows occur as customers send failed items to repair facilities. Lateral shipments can occur between stocking locations.

stocking policy for all parts and location combinations determines the availability of the service-supply chain to meet these demands. This deployment of resources and the rules and procedures for matching supply with demand drives the performance metrics for lead time and customer service.

The interplay between the two hierarchies describes a firm's strategy for managing its service-supply chain (Figure 4). The fastest way to meet targets for responding to customer demands for service is to replace failed products with spare (standby) units positioned at forward locations close to customers. It is also the most expensive way. However, if the products are critical to the customer's operations it may be appropriate. For example, the routers Cisco Systems produces for use with computer servers that support financial transactions at major financial institutions are essential to these operations, and their failures cause major costs. Consequently, Cisco may keep a standby spare close to the customer's facility (perhaps in it) to ensure that downtime is minimal. It may charge the customer a premium for such service. On the other hand, the most economical and the

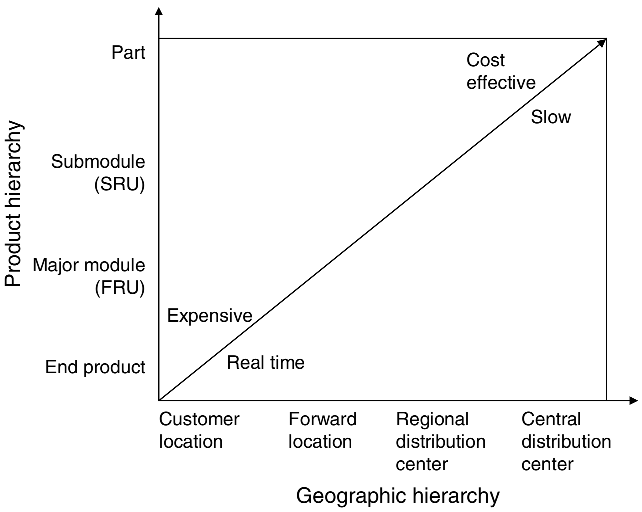

Figure 4: The combination of the asset-deployment decisions across the product and geographical hierarchies determines a firm's service-supply-chain strategy.

slowest way to meet service demands is to identify and replace the failed components of products at a central location.

Service-Asset Decisions

The joint geographic and product hierarchies provide a framework for understanding the choices firms can make for positioning and managing the resources they use to fulfill service entitlements. The geographic hierarchy decisions consider where to locate resources. The product hierarchy decisions consider what to deploy. Service-supply-chain managers must make these decisions for all resource-location combinations. In typical environments, such as Cisco's, they must make millions of such decisions. These deployment decisions are interrelated. An investment in a resource at one location will influence investment decisions for many other item-location combinations. For example, positioning field replaceable units (FRUs) at forward locations can decrease emergency demands for piece parts at higher echelon locations. Similarly, by investing in additional safety stock at a central depot, the firm reduces the lead time for replenishment at the lower echelons connected to it. This lead-time reduction will, in turn, reduce the stocking requirements at the lower echelons. Such decisions are often constrained by the budgets allocated to service organizations. Consequently, assigning a particular asset to one location affects what can be assigned to other locations. Thus, when the budget for asset investment is fixed, the service levels that can be offered to customers at various locations are interrelated; fast service for one customer may imply slow service for another.

In addition to periodically deploying assets, service providers must also manage the flow of such assets over time to replenish stocks and to adjust deployments in anticipation of future service events. They must make tactical decisions concerning replenishment (purchase and repair) of consumed assets and redeploying assets within the service-supply chain (allocating incoming assets, transshipping excess assets, and so forth). To optimize the full set of asset-management decisions, managers must consider such factors as budget, cash flow, and service targets. Finally, they must plan to achieve performance objectives, such as maximizing products' uptime, minimizing total system costs, and minimizing cash-flow requirements for repairs and purchases.

The problem of managing service assets is further complicated by changes in the drivers of resource requirements, such as product-utilization rates and part-failure rates. Thus, the third dimension of resource deployment concerns time. Since demand for assets arises from service events that are highly uncertain, solving the problem of managing assets requires a probabilistic, dynamic representation of the environment.

Service providers must decompose the complex problem of managing service assets into a collection of interrelated decision problems (Figure 5). The basis for the component problems is the period over which to consider managerial trade-offs and objectives (that is, the planning horizon). For the longest planning horizon (months or years), managers make decisions about the design and strategy of their service business. They also may decide on product designs, the service products to offer, and the infrastructure for delivering service. The next level of decision making determines the positioning of strategic assets, such as material, human, and knowledge resources. These decisions require solving an optimization problem periodically (monthly or

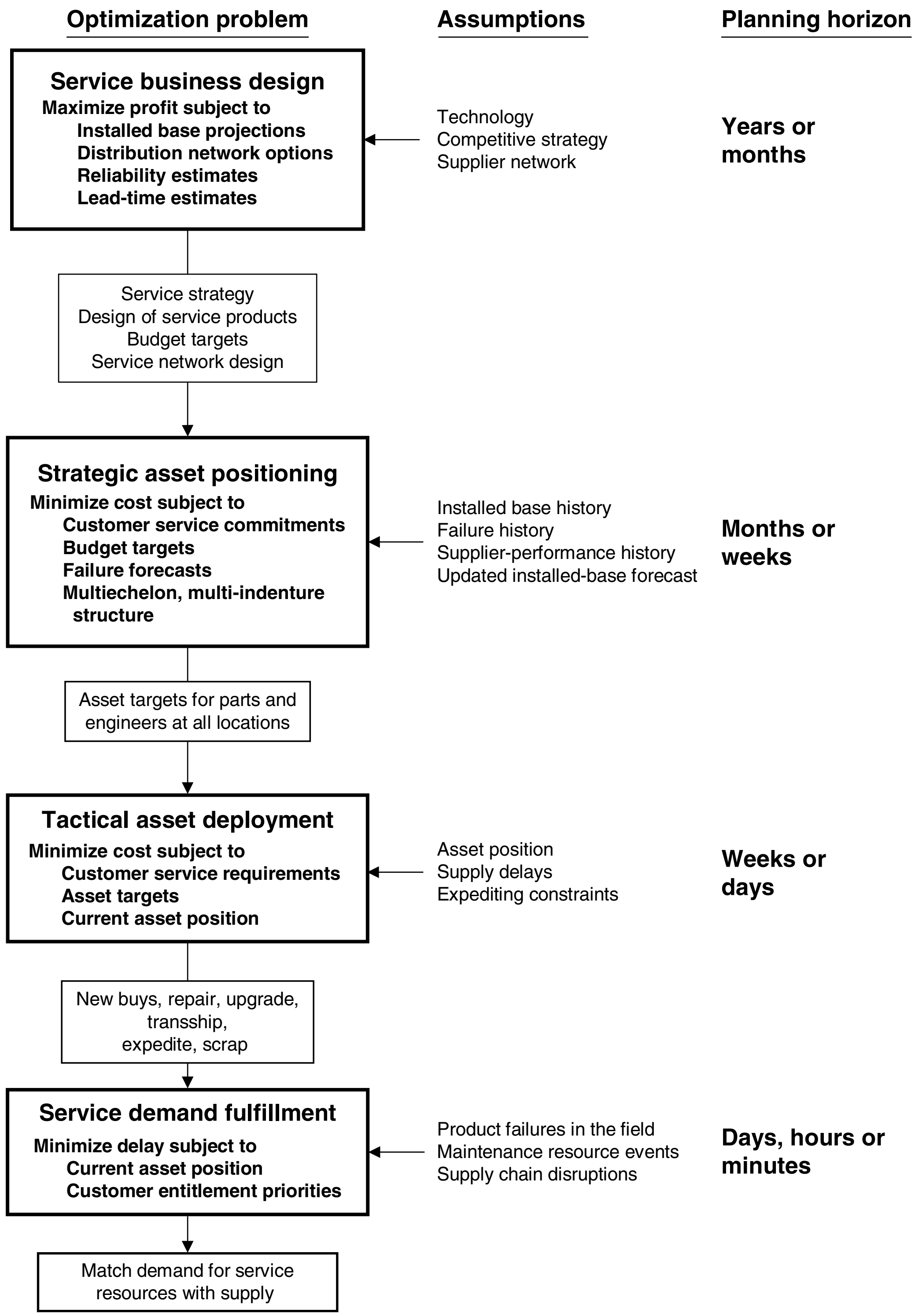

Figure 5: The overall problem of designing the service-supply-chain strategy can be decomposed into separate problems using the chronological planning hierarchy. The solution to each problem provides inputs for the next problem in this hierarchy.

weekly) to implement the DAD strategy. The shortest planning horizon (weekly or daily), incorporates decisions to redeploy assets, for example, material-flow decisions concerning purchases of additional parts, repairs, replenishment, allocation, and transshipment. Finally, decisions to fulfill service demand triggered by product failures, supply chain disruptions, or planned preventive maintenance determine customer satisfaction and operating profitability. Such event-driven decisions occur in real time.

In practice, the frequency with which organizations make decisions in this hierarchy varies, depending on how rapidly the installed customer base changes and how often the firm adjusts its competitive strategy and support capabilities. At Cisco Systems, managers make daily decisions on positioning strategic assets, while at semiconductor equipment manufacturers, such as KLA-Tencor, managers make monthly decisions for positioning assets throughout their service network. In each case, they review asset decisions more frequently in the tactical asset deployment step, for example, every shift or daily, based on constant (near real-time) monitoring of supply and demand conditions.

Managers must make all of the resource decisions for designing the service business, positioning strategic assets, and tactically deploying assets prior to the occurrence of service events that will require those resources. How efficiently firms meet service requirements will depend upon how well they have positioned service resources and assets. Asset positioning is analogous to the purchase of a real option. Contingencies that determine how and where the option to deliver service is exercised include random events, such as product failure, and maintenance-resource consumption. Firms design their service businesses and position and deploy assets based on estimates of future resource requirements along with knowledge of past events that affected the supply and demand of such resources throughout the service-supply chain. The DAD approach recognizes that managers cannot eliminate demand uncertainty through forecasting. They therefore must evaluate trade-offs based on assessments of future risks captured by estimates of the resource-demand probability distribution for specific customer products, locations, and times. The decisions managers make to plan for future events is an exercise in risk management.

Service Demand Fulfillment

Fulfilling service demands (the last mile of after-sales support) is a problem of managing service events after they occur. Managers control the real-time reaction of the service-supply chain to shortages and excesses, determining how it meets the strategic goals of customers. This is where the service product is actually produced. By making intelligent decisions, managers can make the best use of current and projected resource deployments throughout the service-supply chain. Their choices can mitigate the risks of delays from mismatches in resource supply and demand.

Actions to Fulfill Service-Product Demands

When managers face a shortage of assets needed to deal with service events, they can consider various alternatives:

—When material is unavailable at a stocking location, they can transfer the demand to a backup site (usually located at the same or at a higher echelon within the geographical hierarchy) to improve response times and pool risk.

—They can make low-priority customers wait for resources in order to respond to anticipated demand from high-priority customers, thereby rationing scarce resources to ensure availability for high-priority customers. Without such policies, low-priority customers get a free ride with resources maintained for high-priority customers. The low-priority customers can wait until the customer-facing location receives the needed resources, (for example, new parts shipped from a vendor, internal transshipment of excess parts, receipt of a repaired item from a depot, or assignment of a support technician from an alternative location).

—They can restore a failed product by either replacing it with an equal or better product or they can repair the product. The fastest way to repair the product is to replace any defective field replaceable units (FRUs) with working FRUs at the location closest to the customer site. Alternatively they can diagnosis and repair an item from a lower level in the service bill of material at a more central and distant location with the required capabilities and material. They also can substitute superior items that meet the requirements of the product. Typically the substituted items are more expensive and have higher performance quality (for example, a 60 GB disk drive to replace a failed 20 GB disk drive) but lead to more rapid product restoration.

—They can use dynamic pricing and incentives to modify the demands creating shortages. For example, they can provide payments to customers willing to wait for service, as airlines do for overbooked flights. This approach aligns the incentives of the supplier and customer and thus leads to resource-management decisions that optimize total supply chain performance (that is, the first-best solution). However, firms may have difficulty verifying and enforcing contractual terms in such situations.

The heart of the DAD philosophy is to develop asset and resource management strategies to support contingent actions to fulfill service demand. For service-supply-chain managers to react effectively to mismatches between supply and demand, they must plan to be responsive and manage assets accordingly. Such dynamic management of assets sets DAD apart from commonly observed static policies.

For example, consider the rationing policy at a forward (local) location serving two customers who have purchased service contracts covering common spare parts. The firm could maintain separate stockpiles of parts, one for each customer, to avoid potential conflicts. The inventory investment in each stockpile would be commensurate with the service levels agreed to in the contracts for each customer. While this approach is easy to implement and avoids potential conflict and free riding, the service provider can realize risk-pooling gains by serving both customers out of a combined single stockpile. (Cohen et al. 2003; Deshpande et al. 2003 analyze this problem in the context of military spare parts procurement at the US Defense Logistics Agency.) Unless the combined stock is managed appropriately, the customer with the low-service contract will benefit from the inventory required for the high-service contract. For example, if the firm allocates available inventory to customers following a first-come-first-serve rule, in case of shortage situations, the low-service customer can get access to the inventory that should have been reserved for the high-service customer simply because its demand occurred first. This practice denies service to the high-service customer. Managers can design rules to ration inventory that will both avoid such problems and access the benefits from risk pooling if the method they use to determine optimal asset-deployment strategies explicitly plans for this possibility.

Similarly, to effectively apply the substitution policy, which uses a more expensive or higher quality part to fulfill demand for a cheaper or lower quality item, managers must deploy assets in a manner that anticipates the possibility of demand substitution. In effect, the DAD strategy applies the concept of yield management, which has been successful in the hospitality, airline, and rental-car industries, to service delivery.

Thus, if firms do not link post-event actions to pre-event asset-management decisions, the resources they need to meet service-product entitlement targets will not be at the right place at the right time. They will have to make many adjustments to provide customer support, leading to increased emergency-transport, expediting, and shortage penalty costs.

Recourse-based asset planning explicitly accounts for the uncertainty associated with service events and the value of contingency actions and is fundamental to the DAD strategy. Optimization of asset management in the context of a DAD process leads to robust deployments and cost-effective ways of delivering the service promised to customers. In this sense, our approach resembles the real-options framework. Our observations from the field confirm that the dominant mode of thinking prevalent in most companies is based on static, deterministic forecasts that are more appropriate in finished goods DRP or MRP environments.

The DAD strategy differs from current practices in many ways (Figure 6). Asset managers should consider the range in attributes of the items managed, for example, unit cost, expected demand rate, and lead times. Firms often create item-classification groups and use group membership to drive rule-based asset-deployment decisions that result in inefficient decisions. Other traditional approaches to managing assets are based on myopic heuristics, such as looking at one product, one customer, or one location at a time. In some cases, managers decompose the problem into segments, such as a central-depot problem or a field-location-level problem, and solve them independently. Often they ignore the product's

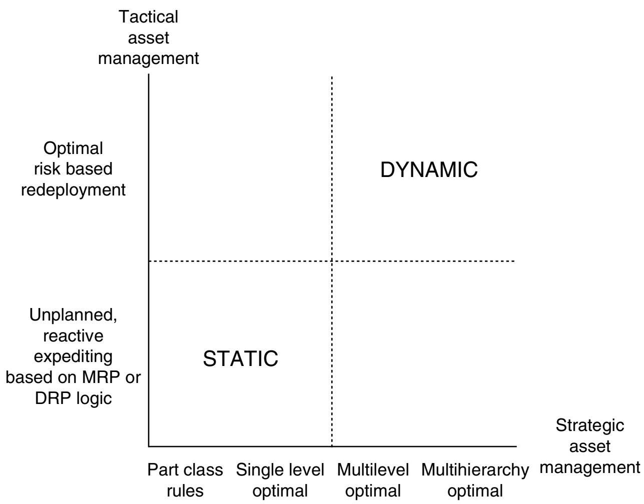

Figure 6: Asset management involves improving the quality of decision making at both the strategic and tactical levels. At the strategy level, a firm can progress from static methods, such as specification of target service levels based on simple part-classification rules or single-location, single-item models to dynamic methods that incorporate trade-offs across multiple locations or indenture levels. The most advanced approach considers simultaneous interactions across both geography and product hierarchies. Tactical-asset decision making begins with static (unplanned reactive expediting) decisions based on MRP or DRP logic (that is, using point forecasts and stationary forecasts) and advances to optimal or dynamic policies based on assessments of changing cost and risk trade-offs. The desired position is in the upper right (dynamic) quadrant. Most firms are positioned in the bottom left (static) quadrant because they do not use the DAD methodology. A firm positioned in the top left quadrant that uses a low level of strategic optimization to position asset investment while managing tactical redeployment of assets in a risk-based manner will achieve limited service benefits or experience high costs. Firms located in the bottom right quadrant may position the right items at the right place but fail to reposition those assets in the short run in recognition of demand changes and anticipated supply shortages. Firms naturally progress from the lower-left to the upper-right quadrant as they move from static asset-management processes to DAD processes.

service bill of material. There is a high degree of interaction across locations and items. In managing assets, considering such interdependencies increases service and lowers costs. We explicitly consider such interactions in our DAD method. Firms must understand these interactions to pool risk across the product and geographical hierarchies in order to realize efficiencies in the service supply chain. For finished goods, separate supply chains may be appropriate for different products. For example, short-life-cycle products are best served by responsive supply chains, while long-life-cycle products require physically efficient supply chains (Fisher 1997). However, in service-supply chains, firms may need the same assets to serve multiple customers, and hence they should engineer as much commonality of resource use as possible. A single service-supply chain, with appropriate allocation rules, therefore is more cost effective in delivering differentiated services than multiple supply chains targeted to different customer types.

If companies optimize the strategic and tactical levels of asset management and effectively integrate them, they will position assets based on long-term strategic planning goals and meet service objectives cost effectively. They will then be able to achieve maximum service from their investment in service resources by redeploying those assets in a manner consistent with strategic-risk-based forecasts and decisions.

Practical Implications of Dynamic Asset Deployment

Many managers of service assets understand the need for flexibility but struggle to accommodate the risks and complexities in providing support services. The DAD model for delivering customer service reflects the way such managers view their efforts to serve customers. It is based on three elements: (1) effective decision-support tools, (2) the information technology needed to support their companies' service goals, and (3) a comprehensive approach to designing service products and to developing a service-business strategy. Firms need a new generation of models and algorithms to deliver differentiated service products dynamically and optimally. These new methods, which are required for implementing DAD, are based on the following steps:

(1) Probabilistic forecasting,
(2) Optimized resource deployment (strategy), and
(3) Optimized resource redeployment and material management (tactics).

(1) Probabilistic Forecasting

Current approaches to supporting decisions about service-supply chains are often based on manufacturing (ERP) or finished-product logistics (DRP) logic. Firms use deterministic forecasts of future demand in deploying resources to schedule production and distribute finished products to the market. For after-sales service, with widely separated customers calling for many low-demand items, the treatment of an estimate of future demand as a point value (deterministic number) can be misleading. As we move from business model to strategy to tactics decisions (Figure 5), the planning horizons compress and uncertainty drops but is not eliminated. In many cases, demand for a part is low at a location (expected use of 0.1 units of a part per month), so the company must make additional assumptions to convert fractional forecasts into discrete values and to schedule the arrival of these rounded quantities.

To support a real-option based model, one must predict future risks accurately. Firms therefore must estimate parameters describing probability distributions of demand by item and location. These forecasts are influenced by historical (time series) and causal factors. Thus a blended forecast, based on a weighted average of time series and causal parameter estimates, should be used. Estimates of the mean time between demands, forecasts of the local installed population, and projected end-product utilization can be used to drive the causal component of such blended forecasts.

(2) Optimized Resource Deployment (Strategy)

To maximize service assets, firms must determine optimal resource deployments for maximizing service performance. At the strategy level, the decision problem can be formulated as a constrained, discrete, stochastic optimization problem that determines the location, quantity, and capabilities of the resources (for example, target inventory-stocking levels for every part-location combination). The objective is to maximize service, to satisfy customer-support needs or to minimize the costs associated with fulfilling those needs. Constraints include interactions in the product and geography hierarchies and limits on budgets, cash flows, and support capacities, probabilistic forecasts of resource requirements, service-agreement entitlements, and constraints on weight and volume. Constraints due to engineering changes and the life cycle of the installed base are also relevant.

(3) Optimized Resource Redeployment and Material Management (Tactics)

At the tactics level, the problem is to manage the flow of materials within lead times. Relevant decisions include new buys, repairs, allocations, excess-stock transshipments, and support staff redeployments. The objective is to minimize both costs and the risk of shortage and delay. The constraints include availability of new and repaired materials, service targets, inventory thresholds, material-flow constraints for normal and emergency shipments, depot capacities, and exception-management thresholds. The constrained optimization problem for this stage of decision support trades off costs against predicted service impact. The firm can also use filtering and triggering mechanisms to limit the need for managers to review the large number of recommended tactical actions that could be generated in this situation (Figure 7).

MCA Solutions, Inc. (www.mcasolutions.com), which specializes in service-supply-chain decision-support tools, has developed a Web-enabled, commercial software platform that incorporates the three steps required for a DAD decision-support system. Its product, the service planning and optimization (SPO) suite, has been implemented in high technology, aerospace, and defense industries where service support is mission critical (Figure 8).

The optimization problems corresponding to the steps in the DAD hierarchy can be formulated as a nonlinear, integer, combinatorial, stochastic, nonstationary model. These formulations are difficult to

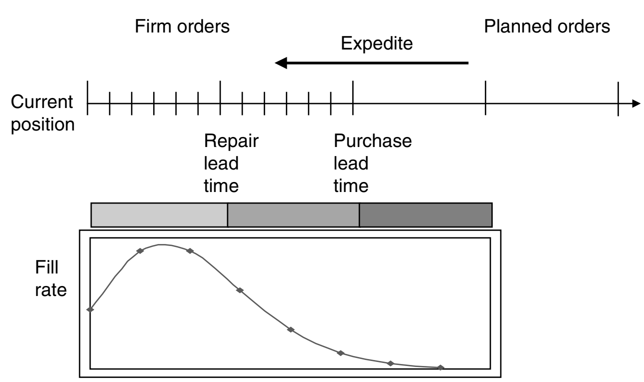

Figure 7: Firms can analyze the tactical problem to quantify a risk profile for each part-location combination based on the part's current inventory position (consisting of the number of units in stock or under repair and en route as returning defectives or generated by confirmed new-buy orders). Managers can use the probability-distribution forecasts for item demand as an input and compare projected material flows against projected demand scenarios. The decision-support system can retrieve the most recent data from enterprise transaction systems.

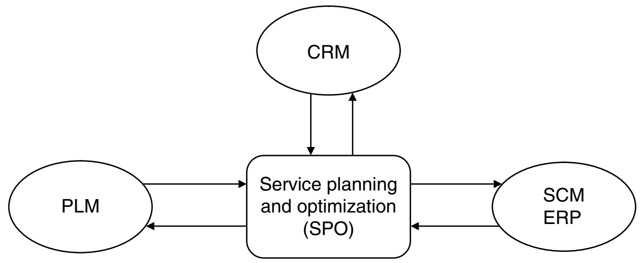

Figure 8: The MCA service planning and optimization (SPO) system is linked to a firm's customer-relationship-management (CRM), product-life-cycle-management (PLM), supply chain management (SCM), and enterprise-resource-planning (ERP) systems. The functions SPO performs include forecasting demand and mean time between failures, positioning inventory in the geographic hierarchy, analyzing service levels, managing service contracts, assessing performance of service contracts, analyzing the effect of design changes on contracts, and allocating, replenishing, and transshipping parts.

analyze and challenging to solve computationally. MCA has developed an approach for dealing with such problems that builds on the extensive research that has been carried out for multiechelon, multi-indenture inventory systems. Its modeling framework and optimization algorithms are based on extensions of methods developed in the military and in high-technology industries. (Sherbrooke (2004) and Muckstadt (2005) review military applications, and Cohen et al. (1990) and Cohen et al. (1999) describe applications at IBM and Teradyne, respectively.) SPO expands the scope of the problems that can be solved and has been implemented as a commercial, off-the-shelf software system to promote its adoption across multiple industries. The steps incorporated in SPO include positioning strategic assets, deploying tactical assets, and fulfilling service demands (Figure 5). SPO uses heuristic solutions that provide near-optimal solutions, as validated through extensive simulation and rigorous testing. The SPO system is effective, fast, and highly scalable, making it practical for large-scale commercial use. Finally, it is one of the few decision-support systems for service-supply chains that has been implemented and empirically tested in a wide range of company settings where it has supported realization of the benefits we predict for DAD.

Conclusion

Firms should adopt dynamic asset deployment (DAD) to design their service products and to manage the deployment and use of service resources. Effective and profitable service delivery depends on managing service assets and fulfilling service demand in a flexible and integrated manner. Indeed, companies must recognize the interactions between these two levels of decision making. To develop a DAD-enabled customer-centric service strategy, we recommend the following:

(1) Firms should recognize that by selling service products they make commitments to customers that the products they have purchased will provide a guaranteed level of value creation. They must design and deliver service products to meet those commitments (for example, equipment uptime) and link the price for service products and the revenues derived from them to actual product performance as realized by the end customer.

(2) Firms should use feedback from field experience to influence product design and to reengineer products so they can deliver service in a cost-effective manner. To do so, they need a deep understanding of how customers derive value from products. Such understanding will lead to designs for serviceability and service-delivery processes that create the maximum customer value (Womack and Jones 2005).

(3) Firms should optimize decisions concerning the capacity, location, and capabilities of the resources they use in fulfilling service demands and account for their impact on cost and service quality. Firms should manage risks by using algorithms that consider cost-service trade-offs and the complex interactions of their decisions throughout the service-supply chain.

(4) Firms should use integrated decision-support tools to maintain service entitlements by linking asset-management and execution systems to a common database that is updated in real time as a result of service-related transactions. Firms also must account for potentially competing priorities for resources.

(5) Firms should invest in processes and information technology to collect, analyze, and disseminate relevant information in a timely and collaborative manner. Recent advances in radio frequency identification device (RFID) technologies and remote sensing and diagnosis capabilities hold great promise for supporting even higher levels of performance.

(6) Firms should design their service-supply chains to achieve their service strategies, determining the

appropriate mix of providers (through outsourcing agreements with third-party logistics and service providers) to support flexible and efficient response mechanisms.

To make good on these recommendations, companies will need to adopt a wholly new paradigm for service-supply-chain management. Companies should treat service delivery as they would a real option, that is, they should invest in resources to purchase the option to deliver service. Subsequently random contingencies occur that determine how they will exercise the option to meet the service requirement.

The concept of flexibility is not new. In fact, it has been a mantra followed by many managers of supply chains for production materials. Indeed, managers of service-supply chains have always recognized that they are engaged in a high-stakes gamble making decisions in a complex and risky environment. Dynamic asset deployment is a way for firms to introduce flexibility and planned responsiveness in delivering service. The DAD strategy and the software tools that support it can help service-supply-chain managers to achieve supply chain flexibility. Companies cannot afford to neglect the potential of this approach in today's hypercompetitive, customer-centric world in which service is often the main competitive differentiator. Companies that have implemented these strategies, such as Cisco Systems, have quickly met with resounding success.

The system Cisco implemented in six months included the decision-support system, SPO. It uses its multiechelon, probabilistic modeling and solution capabilities to dynamically set inventory levels for every part-location combination (over 75 million). Because Cisco handles tens of thousands of service-contract transactions (new or changed) and thousands of service-event fulfillment transactions (material flows) per day, it reoptimizes its asset-deployment decisions daily. It relies on an IT infrastructure that provides real-time visibility into relevant data throughout the service network and on fast solution algorithms.

References

Bijesse, J., M. McCluskey, L. Sodano. 2002. Service lifecycle management (Part 1): The approaches and technologies to build sustainable competitive advantage for service. AMR Research Report, Boston, MA, August, 27.

Cohen, M. A., S. Whang. 1997. Competing in product and service: A product life-cycle model. *Management Sci.* **43**(4) 535–545.

Cohen, M. A., V. Deshpande, K. Donohue. 2003. An empirical study of service differentiation for weapon system service parts. *Oper. Res.* **51**(4) 518–530.

Cohen, M. A., Y. Wang, Y. S. Zheng. 1999. Identifying opportunities for improving Teradyne's service parts logistics system. *Interfaces* **29**(4) 1–18.

Cohen, M. A., H. Lee, C. Cull, D. Willen. 2000. Supply chain innovation: Delivering values in after sales service. *Sloan Management Rev.* **41**(4) 93–101.

Cohen, M. A., A. Tekerian, P. Kamesam, H. Lee, P. Kleindorfer. 1990. OPTIMIZER: IBM's multi-echelon inventory system for managing service logistics. *Interfaces* **20**(1) 65–82.

Deshpande, V., M. A. Cohen, K. Donohue. 2003. A threshold inventory rationing policy for service differentiated demand classes. *Management Sci.* **49**(6) 683–703.

Fisher, M. L. 1997. What is the right supply chain for your product? *Harvard Bus. Rev.* **75**(2) 105–116.

Gallagher, T., M. D. Mitchke, M. C. Rogers. 2005. Profiting from spare parts. *McKinsey Quart.* (February).

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

Sherbrooke, C. C. 2004. *Optimal Inventory Modeling of Systems: Multi-Echelon Techniques*, 2nd ed. Springer, New York.

Womack, J. P., D. T. Jones. 2005. Lean consumption. *Harvard Bus. Rev.* **83**(3) 58–68.

Jim Reily, the vice president of technical support at Cisco Systems, said, "the results of implementing such a dynamic, service differentiated asset management solution have been dramatic for Cisco. Inventory investments have been reduced by nearly 21 percent. This has been accompanied by a simultaneous increase in service level. Working capital requirements have been further optimized by realizing a reduction in the purchase of new parts by almost $65 million. From the customers' point of view, this has resulted in improved productivity, maximized return on network investments, lower operating expenses and increased operating efficiencies."