Interfaces
Vol. 40, No. 5, September–October 2010, pp. 368–384
ISSN 0092-2102 | EISSN 1526-551X | 10 | 4005 | 0368

DOI 10.1287/inte.1100.0507

Hub Group Implements a Suite of OR Tools to Improve Its Operations

Michael F. Gorman
Department of Management Information Systems, Operations Management and Decision Sciences, University of Dayton, Dayton, Ohio 45419, michael.gorman@udayton.edu

Hub Group, a North American intermodal freight rail transportation company, developed an integrated production decision support system to improve its yield management and container allocation. The system combines a suite of analytical tools, including forecasting, error distribution analysis, expected value-based heuristics, and optimization tools to improve the decision making of Hub's frontline employees. It is important to note that the production decision support system was structured to fit within Hub's existing organizational structure, business processes, and information technology systems to reduce project risk, development costs, and operational disruption. In 2008, the system's first full year of use, Hub improved its revenue per load by 3 percent and increased its container velocity by 5 percent. These improvements led to a net return of $11 million in cost savings, which equals 22 times Hub's initial investment during that first year of operation.

Key words: transportation/shipping; decision support systems; freight rail; heuristics; yield management.
History: Published online in *Articles in Advance* July 21, 2010.

Hub Group: Intermodal Transportation

Hub Group is North America's largest intermodal marketing company, with approximately $2 billion per year in revenue. Intermodal freight transportation uses two or more transportation modes, most frequently train and truck. It is the fastest-growing business segment of US freight rail, with over 12 million containers shipped in 2008 (American Association of Railroads 2009), and is second to coal in revenue produced in US freight rail (American Association of Railroads 2009). Figure 1 shows the primary rail ramps in Hub's intermodal rail network and some key origin–destination pairs (i.e., lanes).

A Major Shift in the Industry

During Hub's 37-year history, it has brokered and managed intermodal freight transportation shipments using a fleet of over 150,000 containers and trailers supplied by US railroads. Typical of a nonasset-based intermodal marketing company, Hub's perspective on freight shipment had traditionally been "one way," i.e., based on the profitability of the individual shipment from an origin to a destination. Rail-owned container capacity is shared among intermodal marketing companies, whereas container inventory management is the responsibility of the owning railroad. As Figure 2 shows, asset ownership in intermodal transportation has changed dramatically in the last decade. Quickly shrinking rail-owned trailer and container fleets led Hub to acquire its own fleet of approximately 16,000 containers; becoming an asset-based intermodal marketing company has forced Hub to overhaul its approach. As an integrated transportation network, the company must now manage the balance and utilization of the Hub fleet of containers to avoid wasteful container idle time and repositioning and to assure sufficient container capacity. Further, because of the unique cost structure afforded by the Hub-owned fleet relative to the rail-owned containers, Hub must use them judiciously. Thus, its challenge is to find the right mix of customer traffic and the best allocation of rail-owned and Hub-owned containers to effectively and efficiently serve its customer base.

Load Acceptance, Routing, and the Intermodal Shipment Process

When a shipper requests service from Hub, a Hub customer service representative must determine whether the load is acceptable, taking into account both the load's profitability and the equipment capacity in the shipper's origin region. To provide a quick response to the customer and keep the order, the

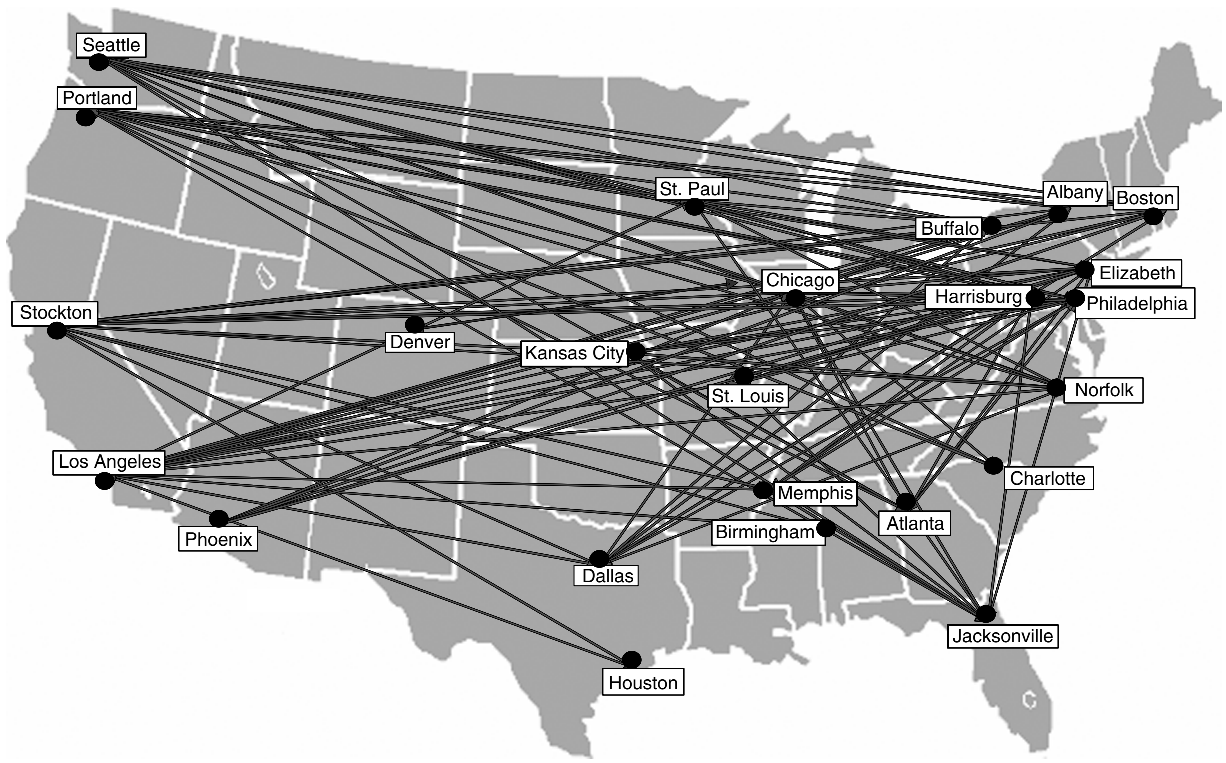

Figure 1: Hub Group's intermodal network provides rail service in lanes between key rail ramp origins and destinations.

representative must make the load accept decision in a timely fashion, usually within a few hours of the initial request. When the load has been accepted, a dispatcher must determine the equipment to use and concomitantly the railroad that will provide the service. Each container fleet can only be used on allowable railroads for that fleet, and each railroad may have a different rate in the same origin ramp–destination ramp city pair; thus, equipment choice affects Hub's cost structure for the shipment.

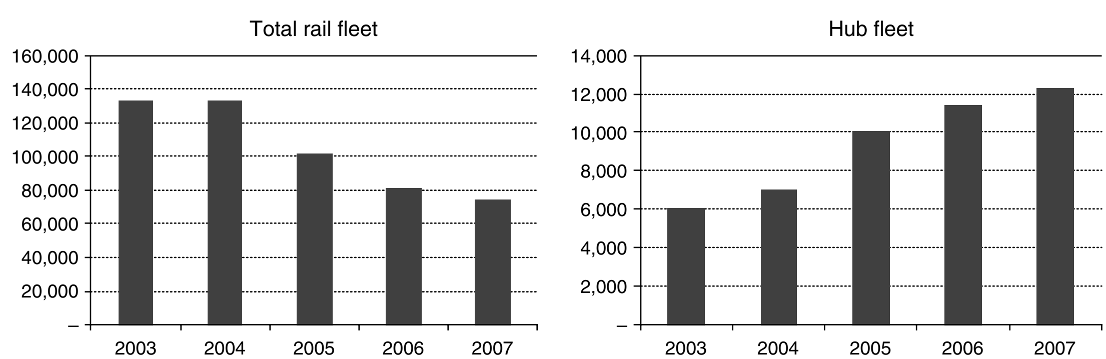

Figure 2: As railroads have reduced their fleets of rail-owned trailers and containers, Hub has increased its fleet of Hub-owned containers, forcing the company to change its approach.

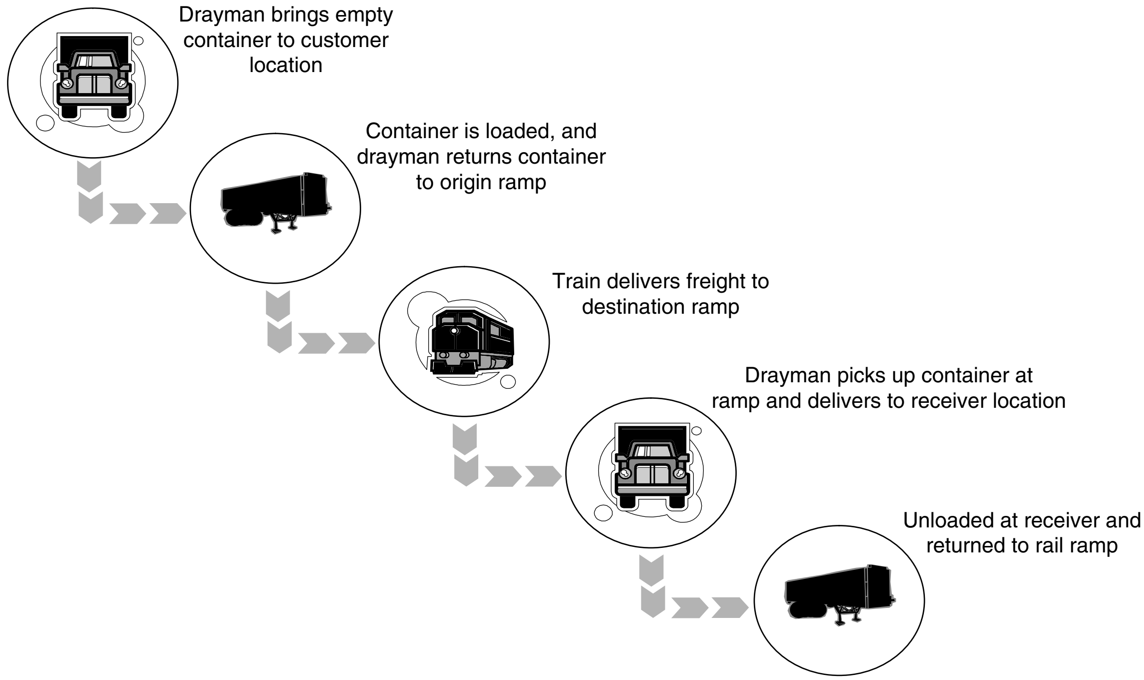

Figure 3: The intermodal process consists of five steps.

Figure 3 shows the five-step intermodal shipment process. First, a drayman delivers an empty container to the shipper. The shipper loads the container, which the drayman then picks up. The drayman moves the loaded container via truck for a relatively short distance from the shipper to the rail ramp (i.e., drayage), where it is transferred to a train and moves via rail for the vast majority of the miles, before finally being drayed from the destination rail ramp to the shipment's receiver. When the container has been unloaded, the customer releases it, and it becomes available for subsequent use. Hub may reuse a rail-owned container for another shipment or simply return it to the appropriate railroad; however, the company must continually manage the disposition of its own fleet by using the container for another load, placing it into storage, or repositioning it to another origin region.

Early Attempts at Improvement

Hub realized that its new fleet would change how it viewed its business. One-way economics would no longer be sufficient to guide decisions; load acceptance and routing decisions on one day represented capacity and profit potential at the load destination a week or two later. Hub needed a new approach to determine how to best deploy its new fleet to maximize its economic advantage. In 2003, Hub and I (henceforth, we) built a strategic Hub fleet allocation model (FAM) to better understand how Hub should allocate the fleet in its network to maximize profitability. As depicted in Figure 4 and described in detail in the appendix, FAM minimizes the cost of serving customer demand for one quarter, subject to a variety of factors: fleet availability constraints, historical transit time, customer equipment return time, Hub fleet balance constraints, and estimated rail fleet availability. Other carrier-specific rebate incentives are also included in the objective function. FAM identified a \$20 million annual profit-enhancement opportunity from an improved Hub fleet deployment, primarily from improved Hub fleet allocation and reduced fleet repositioning.

We developed reports on the optimal mix of Hub fleet and rail fleet containers in each origin–destination pair. However, these benefits were hard to achieve using a strategic model and static reports. FAM did

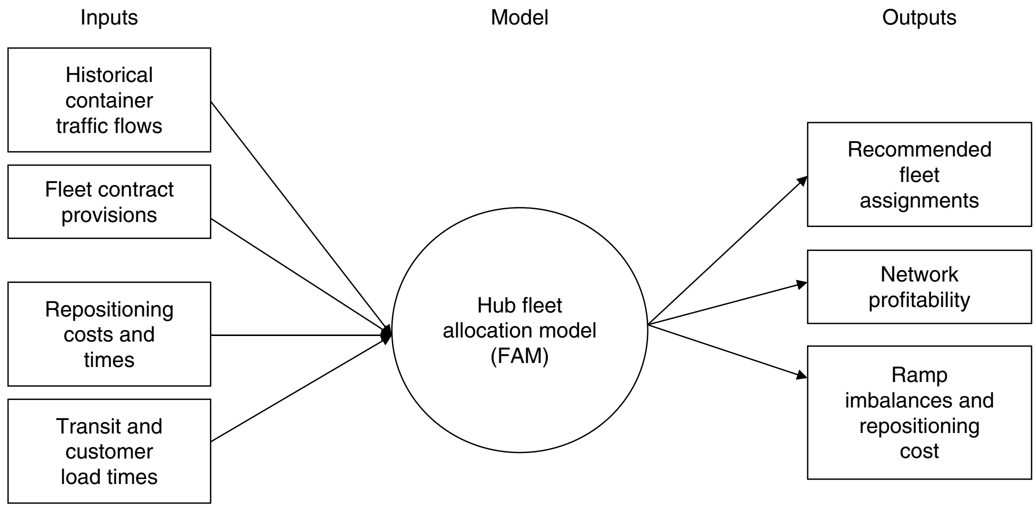

Figure 4: FAM is a strategic model that estimates the benefits of optimization-based Hub container allocation.

not mesh well with the situations that dispatchers faced daily, and dispatchers had trouble translating FAM recommendations into implementable instructions. First, containers become available as they are released from receivers in a highly cyclical day-of-week pattern, which is not reflected in the more long-run, strategic FAM. Second, the recommendations were based on the Hub fleet and rail fleet mix over the span of a quarter, which might not hold on a day-to-day basis. Simple "red light, green light" rules set by managers for use of the Hub fleet in lanes caused dramatic and unprofitable swings in fleet allocation and did not achieve the desired results.

After discussing the options with Hub management and decision makers, we decided that if we were to achieve the benefits that FAM identified, we would have to develop a production decision support system (DSS) that was based on real-time network conditions and provided a better fit for Hub's daily situations and decision processes. We sought to give representatives and dispatchers more holistic information, broader perspectives, and optimization-based recommendations to support their decisions. To accomplish this, we provided Hub with an integrated suite of decision support tools, which allow consistent strategic and real-time decision making to improve the company's networkwide profitability.

To help ensure the project's success, we established several restrictions. To reduce the disruption to the organization, the risk of project failure, and the cost of development, we built the system such that it would not change the organizational responsibilities of representatives and dispatchers or the load-tendering process. The underlying optimization would be "invisible" to the end users. The system also had to fit within the existing information technology systems design. The interface familiar to the end users would be unchanged; new optimization-based recommendations would be based on more sophisticated, optimization-based networkwide economics with the same look and feel as the simpler approach. In short, we had to introduce a network-based view into Hub's existing production information system and one-at-a-time order process.

The Network Optimization Program

In 2007, we implemented a suite of five integrated modules, the network optimization program, to support representatives and dispatchers. Figure 5 depicts the system architecture and decision support modules, which are briefly described below.

- Supply and demand forecasts generate a 14-day rolling-horizon container supply and customer order demand forecast to predict likely container shortages and surpluses.
- The CV model identifies the expected value of a container at a location on a specific date, given the

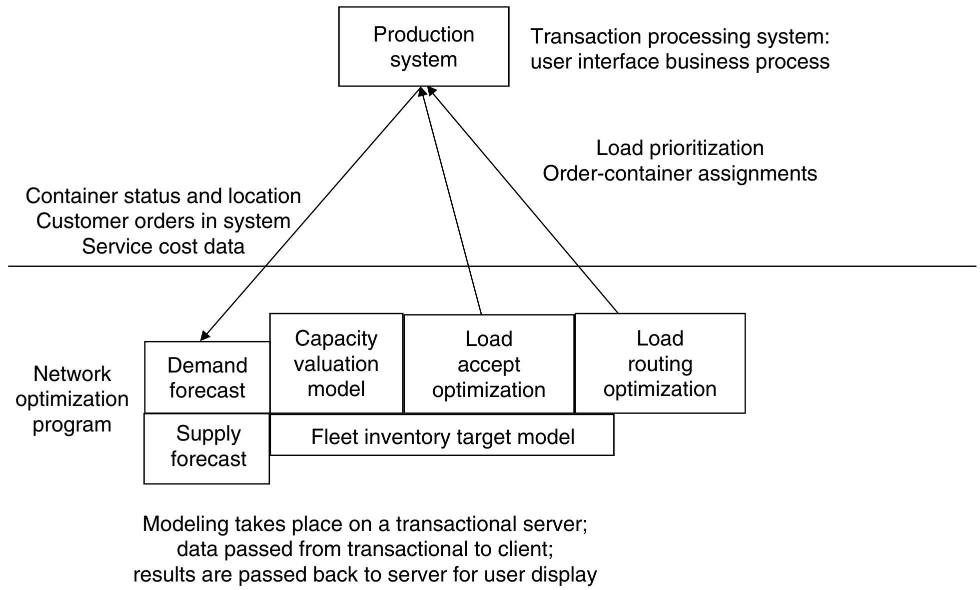

Figure 5: The Hub production system is used by dispatchers and representatives. The network optimization program is an integrated suite of five modules designed to support daily decisions. The client-server architecture has key historical information, which flows from the production system to the optimizer, and provides decision support from the optimizer to the production system.

forecasted surplus or deficit, margin distribution at the location, and distribution of forecast errors.

• The fleet inventory target module establishes a target level of inventory of Hub-owned containers at each location in the network to support day-to-day variations in order levels.

• The load accept optimization (LAO) module sets profitability thresholds for tendered load acceptance, given forecasted demand, anticipated equipment supply, and margin distributions.

• The load routing optimization (LRO) module determines the optimal assignment of accepted loads to the available set of containers and rail service providers.

Supply and Demand Forecasts

Prior to implementing this system, dispatchers and representatives worked without a formal short-term forecast. Supply and demand forecasts, which provided a two-week view of the network, were created daily.

Supply Forecasts. Supply forecasts are based on geographic areas around the 20 major rail ramp-based markets in the Hub network. There are two types of supply: controlled supply and street supply. Controlled supply comprises all containers that Hub customers have unloaded (current supply) or are currently moving under load and will soon have available when the receiver has unloaded the container (future supply). The quantity of future controlled supply is known; however, the timing of future supply availability is uncertain. We used historical unloading-time distributions to predict controlled supply availability. For example, 10 percent of containers are made available the day after delivery, 30 percent in two days, 40 percent in three days, and 20 percent in four days.

Street supply is the number of rail-owned containers that can be attained on a specific date at a ramp location (i.e., "off the street"). This is an expert-based forecast derived from Hub's experience with rail providers and current equipment-supply conditions. The sum of controlled and predicted supply indicates the expected container capacity at a location on a given date.

Demand Forecasts. Forecasts of daily demand are created for approximately 150 major origin–destination pairs in the Hub network; these forecasts constitute 85 percent of the total demand that Hub faces. Smaller shipping lanes are aggregated together by origin so that all customer demand is forecast for each of the 40 major origins in the network. Demand forecasts are statistical forecasts based on customer order behavior. Forecasts are based on traditional time-series methodologies, such as autoregressive integrated moving average and autoregressive estimation, with dummy variables, such as day of week and month of year. Because loaded containers eventually become empty and Hub can reuse them, the long-range supply forecast is based on the projected release dates of forecasted demand.

Table 1 shows a forecasted demand of 20 units for an example lane. We retain the historical error distribution from −5 to 6, and we apply it to the forecast level of 20 to get a nonzero probability forecast range from 15 to 26. The probabilities of each of these demand levels for the lane for that day are based on the historical marginal probability density function. Historical demand forecast error distributions are critical to developing CV estimates for the various levels of container capacity.

Capacity Valuation Model

The CV model estimates the marginal profit potential of a container at a location on a given date. The model creates a CV curve (CVC) used by both the LAO and LRO modules. At an origin, the value is considered an opportunity cost of accepting a load; at a destination, the value is considered the profit potential created by future incremental supply for future moves.

Adelman (2007) discusses the concept of creating profit potential through optimal supply positioning at Hub from a theoretical and strategic perspective; because Adelman's model was not intended for real-time decision support, Hub did not implement it. We created a novel two-step heuristic methodology for estimating the marginal value of incremental container supply to support real-time decision making. The details of the heuristic are given in the appendix.

Step 1. Calculate the marginal expected profitability within a single lane.

The expected profitability of a container depends on the profitability of loads in a location and the probability that the container will be used. The heuristic compares container inventory levels with the expected demand to apply a probability to each unit being used on that date. The CDF is the cumulative density function, which is the sum of all probabilities up to some forecast level, x. We define the inverse cumulative probability density function (InvCDF) for each demand level as the probability

Historical error distribution (u)	Historical probability density function (PDF) (%)	Forecast distribution level ($X = D + u$)	Cumulative distribution function CDF($D \leq X$) (%)	CDF($D \leq X - 1$) (%)	Inverse CDF probability ($D \geq X$) (%)	Marginal expected profit value of capacity $= X$ ($)
−6	0.0	14	0.0	0.0	100.0	400
−5	0.5	15	0.5	0.0	100.0	400
−4	4.5	16	5.0	0.5	99.5	398
−3	8.5	17	13.5	5.0	95.0	380
−2	11.5	18	25.0	13.5	86.5	346
−1	13.5	19	38.5	25.0	75.0	300
0	19.5	20	58.0	38.5	61.5	246
1	13.5	21	71.5	58.0	42.0	168
2	12.5	22	84.0	71.5	28.5	114
3	8.5	23	92.5	84.0	16.0	64
4	5.5	24	98.0	92.5	7.5	30
5	1.5	25	99.5	98.0	2.0	8
6	0.5	26	100.0	99.5	0.5	2
7	0.0	27	100.0	100.0	0.0	0
	Forecasted demand (D) = 20				Profit(Lane) = $400	

Table 1: We applied historical forecast error distributions and probabilities to the forecasted value to obtain marginal and cumulative probabilities. The inverse cumulative distribution function is used to calculate the probability of using a container on a given day. The probability of each forecast level coupled with the load's profitability gives the expected profitability of a marginal container.

of any demand greater than or equal to some inventory level, $x(\text{InvCDF}(x) = 1 - \text{CDF}(x-1))$. Gorman and Ahire (2006) used a similar methodology for replacement-part inventory valuation. We calculate the InvCDF for the error distribution surrounding point forecasts. The InvCDF is applied to the expected profitability at each location to get an expected profitability for container inventory.

From Table 1, the profitability of a container in a lane is \$400. For any number of containers less than 15, the container is worth \$400 because we fully expect to sell at least 14 units. Based on the diminishing probability of its use, the expected profit value of a container falls, as additional containers are considered, to \$246 at the expected demand level of 20 orders (i.e., there is a 50 percent chance of the container being used) and to \$0 if the container supply exceeds 27 units on that day (i.e., there is no chance of selling more than 27 units).

The problem becomes more complicated than that depicted in Table 1 in two ways. First, as noted in Gao and Gorman (2010), origin container capacity is fungible between destinations; the yield potential for containers is based on any number of destinations to which the container might be sent. The expected profitability of a container depends on the number of lanes from an origin and the profitability distribution of the lanes. Second, we note that all customers in an origin–destination lane are not of equal profitability because of their different contractual arrangements with Hub. To represent this complexity, we divided each lane into high-, medium-, and low-profit market segments, in which the median values within each profit-margin class were used as the estimate of profit in each class. For example, if the market segments were represented by even thirds in each lane, roughly the 16th, 50th, and 84th percentiles would be used. Three market segments provided enough distinction between profitability levels in a lane; although we could easily add more, this would cause increased data requirements and complexity with little added ability to assign accurate valuations to capacity.

Table 2 shows an example of CV calculations for two lanes and two market segments. We used the same methodology to calculate the values in Table 2 as we did in Table 1, but we divided each potential level of demand for each lane by the two possible market segments and used the joint probability of each level of demand and market segment to calculate the inverse CDF for each lane.

Step 2. Rank the opportunities by expected profit and calculate their whole-unit equivalent.

Table 3 shows the same data as Table 2; however, we sorted the data by declining marginal expected profitability, interleaving all lanes and market segments for container supply out of the single origin. We took the weighted-average marginal expected profit of fractional units and rounded it to the nearest integer equivalent to get the expected value of an incremental unit of supply on a given date. The CVC is graphed in Figure 6. Given an anticipated level of container inventory for a date, the CVC identifies the CV. For example, a fourth container has an \$86 expected profit value, given the load profitability distribution and the probability of its use.

Hub Fleet Inventory Targets

FAM showed that we needed to identify the appropriate Hub fleet allocation on a long-term basis to maximize the profit potential of the Hub fleet. Because of Hub's railroad contracts, the Hub fleet and rail-owned fleets have distinctly different profit potential. Additionally, for its own fleet, Hub has the responsibility of maintaining its container fleet distribution. FAM's network view does not match the dispatchers' local focus. Therefore, we recast the question: What inventory of Hub containers is desirable at each location to allow a profit-maximizing mix of Hub and rail container fleet assignments?

Conceptualization of the Hub Fleet Cost Function by Origin

Figure 7 converts the global FAM view to an origin-specific inventory cost function. Conceptually speaking, too little Hub fleet inventory might lead to lost business or lower profit levels because of using the higher-cost rail-owned fleet, whereas too much Hub fleet inventory might lead to high idle times, assignment to lower-profit loads, or costly repositioning. We ranked the cost differentials for the Hub fleet from best to worst for each origin in the Hub network to show the trade-offs between the fleets, and we used these differentials for each origin for use as input into the fleet inventory objective function.

	Market segment (k) M1	M2
Profit(OD1) $P(1,1,k)$ ($)	100	500
Prob(Mk | OD1) $\rho(1,1,k)$ (%)	40	60
Profit(OD2) $P(1,2,k)$ ($)	0	800
Prob(Mk | OD2) $\rho(1,2,k)$ (%)	50	50

Origin–destination pair ($i=1$, $j=1,2$)	Demand level (u)	Demand probability density function $\varepsilon(i,j,u)$ (%)
OD1	1	20
OD1	2	50
OD1	3	30
OD2	1	25
OD2	2	45
OD2	3	30

Index (r)	Origin–destination pair (i,j)	Market segment (k)	Market segment probability $\rho(i,j,k)$ (%)	Demand level (u)	Demand level probability $\varepsilon(i,j,u)$ (%)	Joint prob JP$(i,j,u,k) = \rho(i,j,k)\cdot\varepsilon(i,j,u)$	Inverse CDF $1-\text{CDF}(n-1)$	Marginal expected profit (MEP(i,j,u,k)) (%)
1	OD1	M1	40	1	20	8	100	100.00
2	OD1	M2	60	1	20	12	92	460.00
3	OD1	M1	40	2	50	20	80	80.00
4	OD1	M2	60	2	50	30	60	300.00
5	OD1	M1	40	3	30	12	30	30.00
6	OD1	M2	60	3	30	18	18	90.00
7	OD2	M1	50	1	25	13	100	0.00
8	OD2	M2	50	1	25	13	88	700.00
9	OD2	M1	50	2	45	23	75	0.00
10	OD2	M2	50	2	45	23	53	420.00
11	OD2	M1	50	3	30	15	30	0.00
12	OD2	M2	50	3	30	15	15	120.00

Table 2: Demand forecasts for two lanes are split by three market segments. The joint probability of each demand and market segment is applied to the market-segment mean profit to obtain an expected profit for each potential level of demand.

Index (r)	Origin–destination pair (i,j)	Market segment (k)	Demand level (u)	Inverse CDF (%)	Sorted marginal expected profit (MEP(i,j,u,k)) ($)	Expected demand units $=\rho(r)$	Cumulative loads $=\Sigma\rho(r)$	Capacity valuation curve (CVC) ($)	Whole capacity unit
1	OD2	M2	1	88	700.00	0.50	0.50	569.09	1
2	OD1	M2	1	92	460.00	0.60	1.10		
3	OD2	M2	2	53	420.00	0.50	1.60	354.55	2
4	OD1	M2	2	60	300.00	0.60	2.20		
5	OD2	M2	3	15	120.00	0.50	2.70	111.11	3
6	OD1	M1	1	100	100.00	0.40	3.10		
7	OD1	M2	3	18	90.00	0.60	3.70	86.00	4
8	OD1	M1	2	80	80.00	0.40	4.10		
9	OD1	M1	3	30	30.00	0.40	4.50	13.33	5
10	OD2	M1	1	100	0.00	0.50	5.00		
11	OD2	M1	2	75	0.00	0.50	5.50	6.32	6
12	OD2	M1	3	30	0.00	0.50	6.00		

Table 3: By sorting potential demand by marginal expected profitability and combining fractional units into their whole-unit equivalent, we assess the profit value of incremental units of capacity.

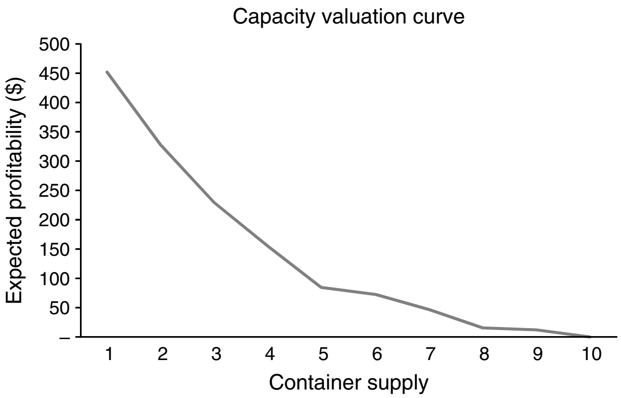

Figure 6: Based on profitability levels, forecast levels, and errors for each lane out of an origin, the CVC describes the relationship of declining expected profitability from expanded container supply on a given date. The LAO module applies actual forecasted container supply level to the CVC to estimate the specific value of capacity on a given date.

Establishing Fleet Inventory Targets

There is a strong weekly pattern of orders and Hub fleet customer releases at each ramp location. Given these patterns, there is an optimum daily mix of Hub fleet that varies greatly from day to day both in total Hub fleet loads and in the percentage of Hub fleet loads. Similarly, there are widely varying target inventory levels by day of week, which are driven by customer order and release patterns. This result varies greatly from what would be implied by FAM; recognizing the day-of-week pattern is critical to achieving a high-profit mix of Hub fleet and rail-owned fleet loads.

To estimate the optimal container inventory targets at each location, we built a nonlinear optimization model based on the regular day-of-week patterns at a location. The decision the distributor faces is whether to use a Hub container on an outbound load, store it for a subsequent day's use (at an inventory carrying cost), or reposition it out of the region. The fleet inventory target model optimizes the net fleet cost advantage in order assignment, subject to the constraints that fleet assignments are less than the fleet inventory, fleet assignments are less than total daily demand, and the following day's inventory equals fleet releases less fleet assignments (see the appendix for modeling details). The result of this optimization is a steady-state allocation of Hub containers to orders and the resulting optimal inventory of Hub containers by day of week. The optimal allocation implies a

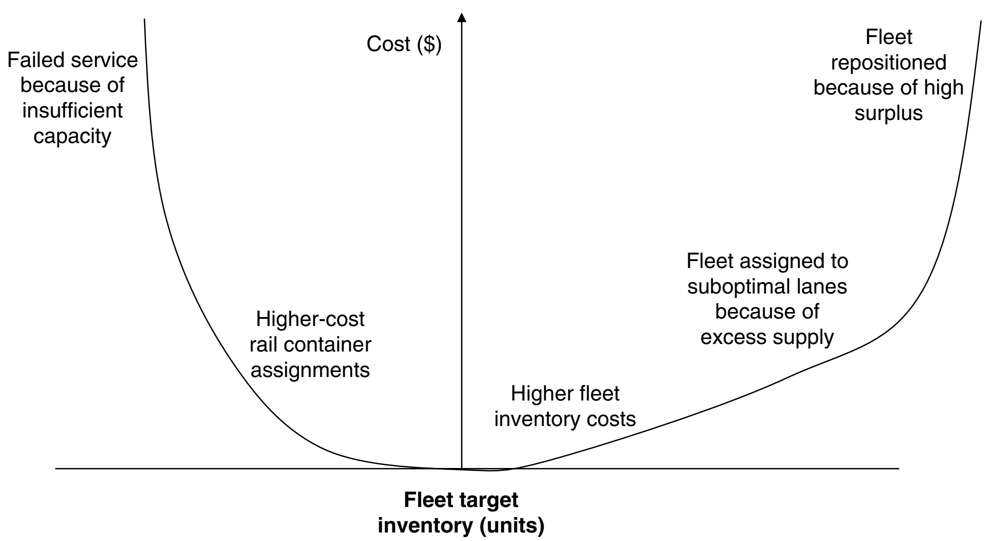

Figure 7: We observe convex costs because of deviations from optimal Hub fleet inventory levels. At the low end, we see service failures and increased costs of using rail container fleets; at the high end, we see higher costs for idle fleet inventory and repositioning.

target inventory level that is used to guide dispatchers' decisions; when the inventory is too high or too low, dispatchers must react.

Tactical, Dynamic or Strategic, Steady-State Modeling Basis?

In setting the fleet inventory target levels, we had to decide between assuming steady state, in which the end-of-week inventory equals the beginning-of-week inventory, and a tactical, dynamic model with a given starting inventory at each location with or without a constraint on the end-of-week inventory level. We noticed an inconsistency with targets based on the two modeling assumptions. We found that in some cases, a given deviation from the optimal steady-state target inventory was fortuitous and resulted in incremental profit potential. In this example, Hub's profits are helped if 30 containers are available on Saturday night, even though it is above the steady-state optimum. In other cases, the opposite might be true. In short, we found the steady-state target inventory model can be misleading when used in a tactical setting.

We ran Monte Carlo simulations of repeated tactical models while varying starting inventory levels for each day of the week and comparing the objective function values in each case. In this way, for any day of the week, we could estimate the true optimum inventory level at each location, which is the level at which the marginal value of incremental inventory is zero in the dynamic setting (see the appendix for details). In our tactical modeling, we were faced with a challenge—how to set the end-of-week conditions. We could constrain the tactical model to achieve the level implied by the long-run, steady-state optimum or take advantage of short-term opportunities, effectively ignoring seven-day-away inventory levels. If we never set a long-run target, dispatchers would always be reacting and not trying to attain strategically sound inventory levels; however, aggressively pursuing a long-run target may obviate some short-term fleet allocation opportunities and create incremental opportunity costs of adjustment to a long-run target. Because there is no clear way to resolve this paradox, additional research on this topic is being conducted (Gorman 2009).

The Implemented Heuristic

Ultimately, we used heuristics based on FAM results to set and update these inventory targets. We translated FAM's optimal equipment allocation in lanes to implied inventory levels, with which dispatchers could more readily identify, at each location. We applied two steps: identify target inventory levels and then identify cost of deviation. Target inventory levels were based on a fixed daily mix of fleet loads and non-fleet loads based on the fleet container mix recommended by FAM. The resulting daily inventory levels from fleet releases and allocation might have to be adjusted to account for possible negative fleet inventory levels or allocations that exceed demands in cases of highly volatile fleet releases and customer orders.

Using FAM for setting inventory targets offered advantages. It integrated strategic and tactical planning tools to create a link between planning and execution, and it considered the entire network when setting inventory targets, rather than a single location at a time. A final practical advantage was that FAM-based allocations and inventory targets were less volatile and easier to understand and manage.

Load Acceptance Optimization (LAO)

Representatives historically had accepted orders with positive one-way profitability on a first-come–first-serve basis. This behavior was especially amplified by the common pool of rail-owned containers. Profit maximization was achieved through volume maximization: intermodal marketing companies that accepted orders and claimed rail-owned container capacity first generated the most volume and profit. With a Hub container fleet and a tightening rail-owned fleet, the representative must now determine whether a tendered load is more desirable than a potential future load it may displace. The objective of LAO is to maximize the expected yield of container capacity. LAO reserves capacity when appropriate for future potential loads and only accepts the loads whose economic value justifies the use of a container.

The LAO module is directly integrated with the production system that the representative uses. The representative must usually make a load accept or reject decision within a few hours because deferral of the decision can reduce customer satisfaction and cause lost orders. We assumed representatives must make decisions immediately on a tendered load. Gao and Gorman (2010) further explore the profit potential for Hub of decision deferral until more orders are received.

To support the sequential representative decision process, we use an expected-value-based heuristic. Supply and demand forecasts and the CVC contribute to the representative's decision to accept a tendered load. A tendered load has expected profitability based on historical profitability of the customer's order plus the CV of the container at the location and future date of the load's destination. Unlike airline yield management, we noted that container inventory can be carried over from day to day; in some cases, customer orders for a date can be "rolled," or served on a later date when there is insufficient capacity on the desired order date. Additionally, there is some uncertainty as to the timing of supply availability. To account for intertemporal rationing and the inherent uncertainty of supply and demand, we take an average CV for seven days after the evaluated date. The seven-day average also adjusts equipment valuations for strong day-of-week effects that are inherent in Hub's customer order and container release patterns. Gao and Gorman (2010) consider capacity rationing and semipersistent inventory effects on intermodal freight yield management more fully.

Container capacity and profitability are determined by the total container supply at a location and the mix of their possible uses (containers can be sent to any destination), not by lane (in contrast to airline yield management in which capacity is fixed by origin–destination pair). Thus, we compare the tendered load's profitability to the "load profitability threshold" for that location and date: the CV of loads out of the customer's origin region plus the CV at the weighted-average destination of loads out of that origin. We accept the load if the tendered load value is greater than or equal to the threshold value.

Customer load profitability and thresholds are updated daily as supply and demand forecasts are updated. The representatives see both profitability values when making a decision; based on their experience, current booking patterns for that day, and their risk-aversion level, they can decide to accept (or reject) a load whose profitability is under the profitability threshold.

Load Routing Optimization (LRO)

Once a load is accepted by the representative, it is turned over to an operations-centric dispatcher who allocates a container to the load. Alternative containers (e.g., Hub-owned or rail-owned) can be assigned to a customer order with little or no difference to the customer, but with substantial potential profit difference to Hub because of the containers' varying cost structures. The dispatcher has some ability to collect orders and make deferred allocation decisions based on more aggregate views of supply and demand; however, loads are typically sequentially dispatched throughout the day while the representative continues to accept new orders, and containers continue to be released from customers. Previously, dispatchers had allocated containers based on a simple greedy heuristic: assign the lowest (one-way) cost container to each load, applied sequentially on a load-by-load basis. This approach had two problems. First, the lowest-cost containers were depleted at each origin first, resulting in expensive and sometimes unprofitable misallocations of containers to subsequently dispatched loads. Second, the future value of the container at the load destination was scarcely considered both because of the lack of forecast information and focus on the one-way economics of each load.

LRO combines forecasts, CVs, and inventory targets to balance near-term costs with future potential profit. LRO minimizes the costs of assigning containers to accepted orders (see the appendix for the formulation). Similar to FAM but in a tactical setting, cost components include the one-way costs of the load (fuel, dray, rail line haul, etc.) and economic penalties for deviation from target inventory levels, less the expected CV of the container at destination. Target inventory levels are a short-horizon proxy for forcing balance in FAM (and were based on FAM modeling). CVs are adjusted by container type in LRO to adjust for different economic profiles of the equipment at various destinations. CVs approximate the effort in FAM to allocate containers to the best lanes; CVs represent the downstream benefit that is fully incorporated in the strategic model. Different from FAM, LRO considers day-of-week effects. If the date of equipment supply differs from the date of customer order, a customer service cost (if the supply date is greater than the order date) or container inventory-holding penalty (if the supply date is less than the order date) is applied.

These costs are minimized subject to the constraints that all orders are assigned equipment, and equipment capacity is not exceeded. Orders include actual accepted orders plus forecasted future orders as placeholders. Container supply includes both Hub-controlled supply (i.e., current available supply and anticipated customer releases from current and forecasted loads) and street supply (supply that can likely be garnered from the rail-owned fleet that Hub currently does not have under load). To assure feasibility in each model run, phantom supply is available at high cost to the model and is interpreted as "no supply."

The costing of each possible container assignment is carried out prior to optimization in preprocessing, so the form of the objective function is linear. The problem is a transportation problem, but the forecasted supply and demand values can take on non-integer values. Of course, dispatchers only dispatch equipment to actual orders, not to forecasted demand; they apply actual, not forecasted, supply to these orders. Thus, the relevant model recommendations used by dispatchers are integer. Supply and demand forecasts are important primarily as "placeholders" for future assignments, so that model solutions are not near-sighted.

The implementation of LRO in a transactional environment and system was a challenge. The dispatcher must contact and get confirmation from the drayman (who handles empty container pickup and delivery) for the dispatch. Because of time constraints on the dispatcher, latent or incorrect data in the model, or drayman availability, the model's recommended assignment is not always easy to follow. Consistent with the previous dispatcher process and system information, LRO provides the dispatcher with a ranked list of recommended assignments. Where the list was previously based on one-way economics, now we sort the recommendations based on the dual values from the optimization to capture the network cost of choosing an option other than the optimum. The difference in dual values allows the dispatcher to understand the cost of deviation from optimum; in many cases, the difference between first, second, and third options is minimal, and the dispatcher can work with the drayman on convenient solutions with small incremental network cost. In the unusual case in which LRO uses fractional units of forecasted supply to meet an order, both options for meeting the order are presented to the dispatcher as "best," and the model is indifferent to which is chosen. (This case is unusual because of the service penalty associated with future forecasted supply and current actual orders.)

The model is resolved every 10 minutes throughout the day to account for new orders and container releases and for previously made routing decisions of dispatchers. Thus, deviations from optimum and changing conditions are reflected in subsequent model recommendations. Frequent resolving of LRO keeps recommendations up-to-date. Gorman et al. (2010) present a similar implementation approach in railcar distribution.

Best Practices

It was important for us to minimize disruption and risk in Hub's first experiment with an operations research (OR)-based production DSS. Although integrating networkwide decision technology within a transactional production operating system was a challenge, it was important to keep the project manageable. We also made sure we understood the organization, business process, and system limitations before trying to implement the new system.

The strategic FAM model served as a proof of concept, which helped to justify the more ambitious real-time DSS. Although the benefits were not achieved with the strategic model, the opportunity was identified, and key observations were made for the production model, including setting target inventories by location rather than target mixes by lane.

Challenges

The transition to advanced decision support had its challenges. First, when simple one-way economics are augmented with target inventory deviation costs and future probability-based profit-potential values, the network-based values presented to representatives and dispatchers are not as familiar or as predictable as they were previously. Model compliance is an issue when dispatchers override model recommendations because they are inconsistent with the lowest-cost, one-way container choice that they see as more intuitive. Second, when faced with such situations, the more complicated CVs and network dual values are more difficult to audit, diagnose, and understand. Thus, when an unintuitive CV is presented to the user, it can be difficult to explain and diagnose. Improved and expedited audit-reporting capabilities are necessary to both boost confidence in model results and improve model compliance. Third, some dispatches occur before the LRO model can evaluate them. For example, large customers have container pools on site and can place an order and assign equipment simultaneously. Because dispatchers cannot affect these assignments, the model applicability is hampered. Finally, we note that LAO and LRO are based on the assumption of allocating scarce container resources among competing uses. To the extent that economic downturns create excess capacity in the rail-owned and Hub container supply, forward-looking and networkwide-based model recommendations become less appropriate, and simply choosing the lowest-cost, one-way container achieves good results.

Business Impact

These models have a dramatic impact on Hub's operating philosophy. They provide frontline decision makers with the ability to make choices based on a broad perspective of the Hub network and over a two-week horizon rather than on a one-load-at-a-time, one-way cost view.

The system has made the frontline workers more effective at improving Hub's profitability. In 2008, Hub realized an average of $22 per load from improved selection of rail routes because of LRO. Hub has improved its ability to accept loads that have high profit value, as well as value in the future where the load is terminating. Although the exact dollar savings are difficult to estimate, Hub noted a 3-percentage-point drop in low-value loads handled in 2008. Finally, in 2008, Hub realized over a 5 percent improvement in container velocity resulting from better positioning of equipment for the next load; to some degree, LRO and LAO enabled this.

Hub's performance improvements can also be seen in other ways. In 2008, Hub enjoyed improved equipment velocity throughout its network from less time in inventory, reductions in the magnitude of equipment surpluses and deficits, and reduced empty repositioning to avoid equipment surpluses. In 2008, Hub estimated that its benefit from these efforts was $11 million per year on an expenditure of only $500,000—a dramatic return of 22 times its investment per year.

Summary

Hub Group has taken a major step toward improved container management through the implementation of a suite of integrated optimization-based tools to support the daily decision making of customer service representatives and dispatchers. To minimize project risk, operational disruption, and development costs, we designed the system to function within the existing organization and process.

Supply and demand forecasts project container capacity excesses and shortfalls. Forecast error distributions and load profitability predict the value of regional container capacity over that horizon and are used to accept or reject tendered loads to maximize the yield of the containers. Finally, container capacity of different cost structures are allocated among accepted loads to minimize short-term costs and maximize future potential profitability. In its first year using this system, Hub realized an $11 million benefit in improved performance, among other benefits.

From a research perspective, this application has led to the development of a novel capacity-valuation heuristic that can be used in the unique environment for yield management in intermodal freight transportation. The application has also opened new avenues for theoretical research in intermodal freight yield management (Gao and Gorman 2010) and evaluation of strategic, steady-state modeling and dynamic, tactical modeling paradigms for modeling target inventory capacity in freight transportation systems (Gorman 2009).

Appendix

Sets used for all models

O set of all origins, indexed by i.
D set of all destinations, indexed by j.
C set of all container types, indexed by c.
K set of all market segments, indexed by k.
T set of days in the planning horizon, indexed by t.

FAM Formulation

Parameters

- C_{ijc} cost of assigning container type c to a load in lane (i, j).
- RC_{ijc} container repositioning cost for container type c in lane (i, j).
- S_c total container days of supply for Hub containers of type c.
- I_{ic} maximum net negative imbalance (net supply) for rail-owned containers c at origin i.
- T_c minimum threshold for rail-owned container type c to earn rebate.
- t_{ij} transit time for loads assignment or repositioning in lane (i, j).
- MU_c a reasonable upper bound on volumes in container c.
- MD_c a reasonable lower bound on volumes in container c.
- B_c dollar rebate per load earned if the volume in container type c, V_c, is greater than 0.
- D_{ij} level of customer orders for all lanes (i, j) over a quarter.

Variables

- x_{ijc} allocation of container type c to origin–destination pair (i, j).
- V_c difference of the volume of loads in container type c and the rebate threshold T_c.
- U_c equal to V_c if V_c is positive; $D_c = |V_c|$ if V_c is negative.
- r_{ijc} repositioning volume of container type c to maintain conservation of container flow in lane (i, j).
- z_c, y_c binary indicator variables: $z_c = 1$ when $V_c \le T_c$, $y_c = 1$ otherwise.

$$\text{minimize cost } \sum_{i\in O}\sum_{j\in D}\sum_{c\in C} C_{ijc}x_{ijc} + \mathrm{RC}_{ijc}r_{ijc} - \sum_{c\in C} B_c U_c \tag{1}$$

subject to

$$\sum_{c\in C} x_{ijc} = D_{ij} \quad \text{for all lanes } (i, j), \tag{2}$$

$$\sum_{i\in O}\sum_{j\in D}(x_{ijc}t_{ij} + r_{ijc}t_{ij}) \le S_c \quad \text{for all containers, } c \text{ in } C, \tag{3}$$

$$\sum_{j\in D} x_{ijc} + \sum_{j\in D} r_{ijc} = \sum_{j\in D} x_{jic} \quad \text{for all } c \text{ in } C,\ \text{for all } i \text{ in } O, \tag{4}$$

$$\sum_{j\in D} x_{ijc} - x_{jic} \le I_{ic} \quad \text{for all } c \text{ in } C,\ \text{for all } i \text{ in } O, \tag{5}$$

$$V_c = \sum_{i\in O}\sum_{j\in D} x_{ijc} - T_c \quad \text{for all } c \text{ in } C, \tag{6}$$

$$V_c = U_c - D_c \quad \text{for all } c \text{ in } C, \tag{7}$$

$$0 \le U_c \le \mathrm{MU}_c z_c \quad \text{for all } c \text{ in } C, \tag{8}$$

$$0 \le D_c \le \mathrm{MD}_c y_c \quad \text{for all } c \text{ in } C, \tag{9}$$

$$z_c + y_c \le 1 \quad \text{for all } c \text{ in } C, \tag{10}$$

$$y_c, z_c \varepsilon \{0, 1\} \quad \text{for all } c \text{ in } C. \tag{11}$$

The objective function (1) is to minimize the sum of container assignment and repositioning costs C_{ijc} and RC_{ijc}, respectively. All demand volume requirements D_{ij} must be met (Equation (2)), the total container days of supply for Hub containers type c, S_c, cannot be exceeded by assignment and repositioning time for loads and repositioning in lanes t_{ij} (Equation (3)), and the balance of container flow at each location in the network must be preserved via load assignments or the decision to undertake repositioning r_{ijc} (Equation (4)). There is a maximum net supply of rail-owned containers I_{ic} for originated loads out of an origin (Equation (5)). If volume minimum threshold T_c is exceeded, an equipment-specific rebate, B_c, per unit is earned (Equations (6)–(11)).

CV Model: CVC Derivation Heuristic

Forecast Parameters

- F_{ij} forecasted levels (expected values) of demand for all lanes (i, j).
- P_{ijk} profit level for lane (i, j) and profit class k.
- e_{iju} daily forecast error level, u, of forecast F_{ij} for each lane (i, j), $u = 1 \ldots U$.

ε_{iju} marginal probability of demand level $u = F_{ij} + e_{iju}$, in lane (i, j).
ρ_{ijk} probability of any forecasted demand being from profit level P_{ijk}.

Calculated Values

JP_{ijuk} joint probability of any forecast level u in lane (i, j) and market class k; $\mathrm{JP}_{ijuk} = \rho_{ijk}\varepsilon_{iju}$.
CDF_{ijuk} cumulative distribution function for errors u in lane (i, j) and profit class k.
InvCDF_{ijuk} inverse cumulative distribution function = the complement to CDF_{ijuk}.
MEP_{ijuk} the marginal expected profit in lane (i, j), for demand level u, for profit class k.

Our objective is to maximize the expected profit for accepted loads at a given location i by accepting the most profitable loads from lanes (i, j) and market segments K. The heuristic follows two steps: (1) calculate MEP_{ijuk}, based on InvCDF_{ijuk}, for each possible forecasted level; and (2) rank the opportunities by expected profit and calculate their whole-unit equivalent, which generates the CVC for whole-capacity units based on a weighted average of market segments.

Step 1. Calculate the marginal expected profit in a lane (see Table 2) (note that in these steps, "!" denotes a comment).

```
For all i in O
      ! all destinations from the origin
    For all j in D
    CDF = 0
          ! all forecast levels for the lane
    For all u = F_ij + e_ij1 to F_ij + e_ijU
        For all k in C    ! all market segments
          InvCDF_ijuk = 1 − CDF_ijuk
            ! marginal expected profit
          MEP_ijuk = InvCDF_ijuk P_ijk
          JP_ijuk = ρ_ijk ε_iju    ! joint probability
          CDF = CDF + JP_ijuk
        Next k    ! market segment
      Next u    ! forecast level
    Next j    ! destination from origin
  Next i
End.
```

Step 2. Rank by expected profit and calculate their whole-unit equivalent (see Table 3).

```
For all i in O,
    Sort all data by MEP_ijuk.
      ! from lowest to highest MEP in each origin
    Create index r, on all arrays
    CL = 0
      ! all destinations, market classes, forecast
      forecast levels
for r = 1 to j · c · U
    Unit = 0
            ! haven't crossed whole unit of capacity
      Do while INT(CL) = INT(CL + Unit)
        Unit = Unit + ρ_r
        EP = EP + MEP_r ρ_r    ! weighted sum
      End do
        ! Nearest whole-capacity unit
      CapUnit = INT(CL + Unit)
        ! weighted avg; full-unit equivalent
      CVC(CapUnit) = EP/Unit
      CL = CL + Unit    ! next whole-capacity unit
    Next r
End.
```

Fleet Inventory Target Optimization and Implemented Heuristic

Parameters

L_t the total possible loads on day t.
HFB_t Hub fleet benefit for day t—estimated statistically based on the relative Hub fleet advantage over rail.
x, y, and z statistically estimated Hub fleet profit parameters based on history.
FR_t fleet releases from customers each day t.
I_7^s constrained value of strategic optimum inventory on day 7 of the week (optional).

Variables

FL_t the number of fleet loads to assign each day t.
I_t fleet inventory on each day t.

$$\text{maximize } \mathrm{OBJ}^* = \mathrm{HFB}_t - \sum_{t=1}^{7} \mathrm{HC} \cdot I_t \tag{12}$$

subject to

$$\mathrm{HFB}_t = x(y - \mathrm{FL}_t/L_t)^z \quad \text{for } t = 1 \ldots 7, \tag{13}$$

$$0 \le \mathrm{FL}_t \le L_t \quad \text{for } t = 1 \ldots 7, \tag{14}$$

$$I_{t+1} = \mathrm{FR}_{t+1} - \mathrm{FL}_{t+1} + I_t \quad \text{for } t = 1\ldots 6, \tag{15}$$

$$I_t \geq 0 \quad \text{for } t = 1\ldots 7. \tag{16}$$

The objective (Equation (12)) maximizes the fleet profitability relative to rail-owned fleets. The fleet benefit is a nonlinear function of its intensity of use (Equation (13)), where fleet loads are some proportion of total loads (Equation (14)). Hub fleet inventory dynamic equation (Equation (15)) and nonnegativity conditions (Equation (16)) assure feasible Hub fleet inventory levels. Starting and ending inventory conditions depend on the modeling paradigm. Strategic (steady-state) and tactical (dynamic) optimization were evaluated to set targets. In the strategic model, starting inventory is governed by ending inventory to assure repeatability (Equation (17)):

$$I_1 = I_7 + \mathrm{FR}_1 - \mathrm{FL}_1. \tag{17}$$

In the tactical model, starting and ending inventories are governed by Equations (18) and (19):

$$I_1 = \text{some starting value}, I_0, \tag{18}$$

$$I_7 = I_7^s \text{ (recovery model to strategic optimum)}. \tag{19}$$

(Equation (19) is omitted and ending inventory is unconstrained in reactionary mode.)

To determine the cost of deviating from target levels, we solve both models repeatedly in a Monte Carlo simulation for each day of the week t and constrain I_t for all levels of inventory to evaluate the change in cost from the optimal level OBJ* from strategic or tactical modeling. The decision variable, FL_t, simultaneously determines the target fleet inventory levels, our output of interest for use in LRO.

Load Routing Optimization (LRO)

Parameters

- C_{ijc} the direct cost of assignment of container c in lane (i, j).
- TI_{ict}, TI_{jct} target inventory deviation costs at origin i and destination j, respectively, for container type c on day t.
- CV_{jct} capacity value created at the destination location j for container type c on day t.
- SV_{jct} service penalties mismatch of supply and demand at location j for container c on day t.
- Inv_{jct} inventory cost at location j for container type c on day t.
- $D_{ijt} =$ customer orders, or demand, in lane (i, j) on day t.
- $S_{ict} =$ supply of containers, c, at origin i on day t.

Variables

x_{ijc} assignment of container type c in lane (i, j).

The LRO formulation follows.

$$\text{minimize cost} \sum_{i\in O}\sum_{j\in D}\sum_{c\in C}\sum_{t\in T}(C_{ijc} + \mathrm{TI}_{ict} + \mathrm{TI}_{jct} - \mathrm{CV}_{jct} + \mathrm{Inv}_{jct} + \mathrm{SV}_{jct})x_{ijct} \tag{20}$$

subject to

$$\sum_{c\in C}\sum_{t\in T} x_{ijct} = D_{ijt} \quad \text{for all } (i, j), \tag{21}$$

$$\sum_{j\in D} x_{ijct} \leq S_{ict} \quad \text{for all } c, \text{ for all } i, \text{ for all } t. \tag{22}$$

Equation (20) minimizes the direct cost of assignment C_{ijc} plus deviation from target inventory costs at origin and destination, TI_{ict} and TI_{jct}, respectively, from fleet inventory targets, and inventory and service costs, Inv_{jct} and SV_{jct}, respectively, less the equipment-adjusted capacity value created at the destination location CV_{jct} from the CV model (profit potential is treated as a contra-cost). Constraints assure all demand is met (Equation (21)), and supply in each day is not exceeded (22). Supply and demand that do not share the same time period can be paired; the timing of S and D is accounted for via service penalties (SV_{jct}) that apply when t of $S_{ct} > t$ of D_{ijt}. Conversely, inventory costs (Inv_{jct}) apply when t of $S_{ct} < t$ of D_{ijt}.

Acknowledgments

I would like to acknowledge the significant contributions of Gregg Grabijas, Chris Kravas, Jason Letchford, and numerous other Hub Group employees who contributed to the design, development, implementation, and ongoing management of this project, and without whose help this work would not have been possible. I would also like to acknowledge the support of my colleagues at the University of Dayton School of Business.

References

Adelman, D. 2007. Price-directed control of a closed logistics queueing network. *Oper. Res.* **55**(6) 1022–1038.

American Association of Railroads. 2009. Rail intermodal keeps America moving. Retrieved August 31, 2009, http://www.aar.org/InCongress/~/media/AAR/BackgroundPapers/Intermodal%20Nov%202009.ashx.

Gao, L., M. F. Gorman. 2010. Optimal load tendering and acceptance policies for intermodal freight transportation. Working paper, University of California, Riverside.

Gorman, M. F. 2009. The consistency of strategic and tactical modeling of rolling stock inventory targets. Working paper, University of Dayton, Dayton, OH.

Gorman, M. F., S. Ahire. 2006. A major appliance manufacturer rethinks its inventory policies for service vehicles. *Interfaces* **36**(5) 407–419.

Gorman, M. F., D. Acharya, D. Sellers. 2010. CSX Railway uses OR to cash in on optimized equipment distribution. *Interfaces* **40**(1) 5–16.