Santa Fe Railway Uses an Operating-Plan Model to Improve Its Service Design

Michael F. Gorman
Burlington Northern Santa Fe Corporation
3017 Lou Menk Drive
Fort Worth, Texas 76131

Santa Fe Railway's intermodal business unit has developed an innovative approach to designing its service offerings. The operating-plan model (OPM) minimizes the schedule-related costs of service subject to rail-operating capabilities while meeting customers' expectations for service. The algorithm produces a weekly train timetable and assigns traffic to trains. Because of the problem's size and complexity, a combination of genetic and tabu searches is used to search for successively better operating plans. The OPM shows the potential to improve global service by four percent while reducing costs by six percent over the existing operating plan. Santa Fe Intermodal has realized savings by applying the OPM to more narrowly focused problems.

Santa Fe Railway faces increasing demands for customer service, cost pressures, and changing market conditions. The new and fast-paced intermodal business area, in which traffic moves on some combination of ship or truck and train, faces a particularly strong challenge. Santa Fe has averaged almost eight-percent growth per year in intermodal traffic handled since the intermodal business unit's inception in 1989 to an estimated 1996 level of over 1.6 million units per year (Figure 1).

Santa Fe Intermodal is exploring innovative approaches to creating its intermodal service design, which is the process of

0092-2102/98/2804/0001/$5.00
This paper was refereed.

TRANSPORTATION—FREIGHT/MATERIALS HANDLING
TRANSPORTATION—SCHEDULING—VEHICLES

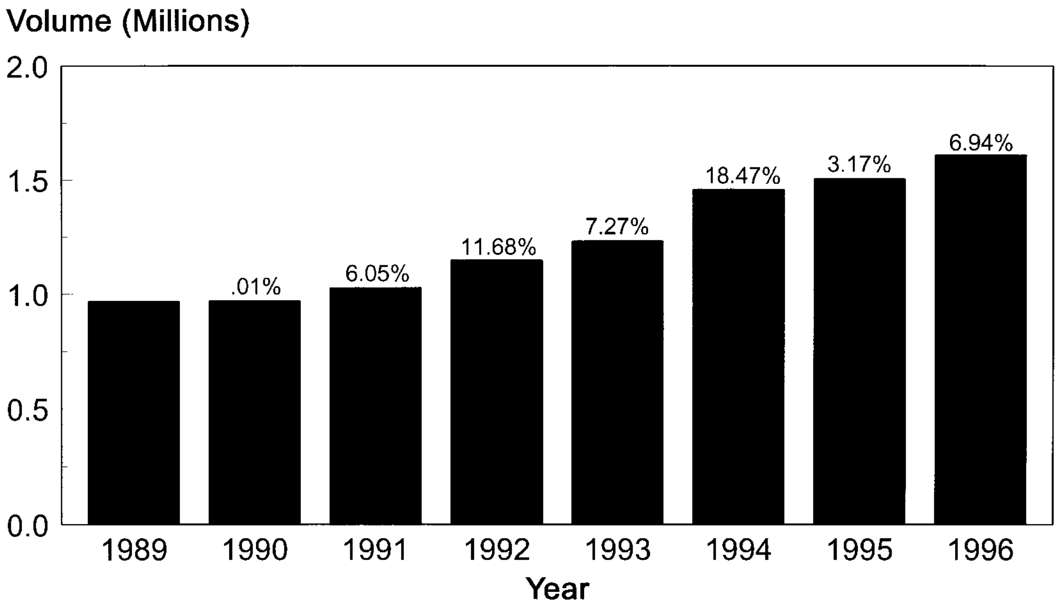

Figure 1: Santa Fe intermodal volumes have grown at almost eight percent per year since its inception in 1989.

developing and modifying a railroad's operating plan. The operating plan of a freight railroad is the weekly train timetable and traffic assignment strategy that defines the service it offers to customers and its daily operations. The operating plan is essential to the smooth and profitable operation of the railroad. An effective plan strikes a balance between customer service and operating efficiency.

Historically, service design in freight rail has been rather myopic. Service designers adjust existing operating plans to gain incremental refinements in current operations. They generate ideas for improvements based on their past experience with train schedules and their expertise and creativity. However, when they try to accommodate changing traffic patterns, service designers find it difficult to evaluate the networkwide ramifications of major schedule revisions based solely on historical scheduling information and their general intuition about rail operations. To keep up with the rapidly changing needs of its customers and to reduce operating costs, Santa Fe needed a tool to help it identify opportunities for sweeping adjustments to the operating plan.

The operating-plan model (OPM) builds the operating plan for the intermodal business unit from scratch, helping the railroad to develop a plan best suited to current and anticipated traffic patterns that is not constrained by traditional thinking or historical schedules. For input, the OPM needs only customer traffic patterns and service requirements, rail operating capabilities, and operating costs (Figure 2). It creates a detailed operating plan that includes train schedules, traffic-to-train assignments, and yard and line train activity.

The OPM is a strategic planning tool designed for long-term planning and lasting

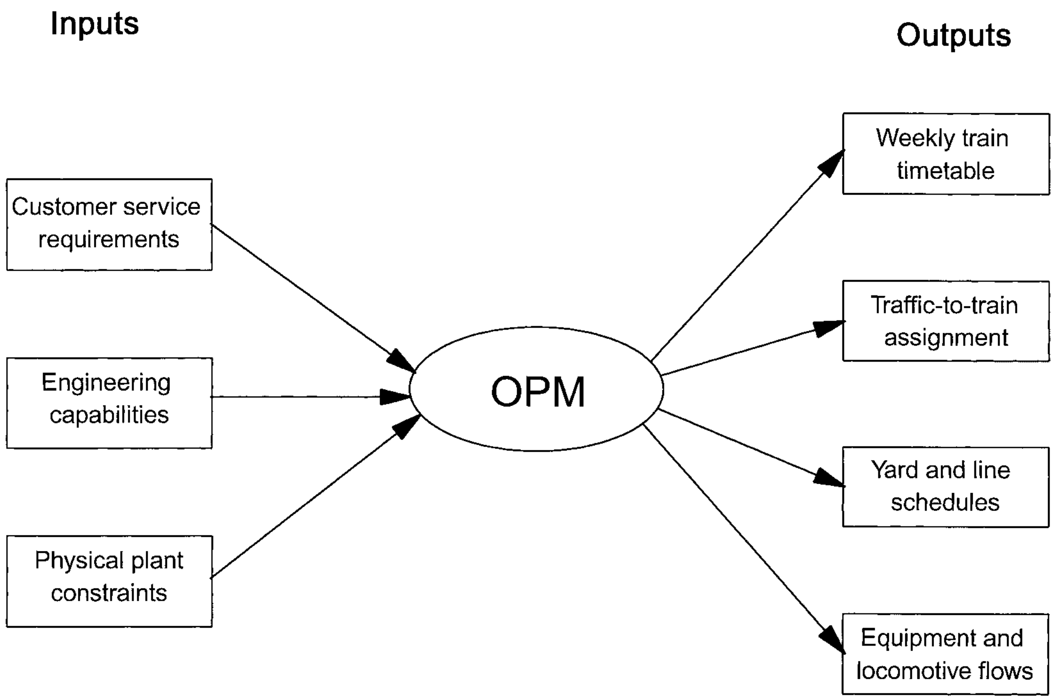

Figure 2: The operating-plan model (OPM) takes basic customer, engineering, and physical-plant data and creates a detailed operating plan with train, traffic-assignment, yard, line, and equipment information.

structural changes to the operating plan. Although it is particularly well suited for making broad-based changes to the operating plan, it has also been successfully used to answer more narrowly focused and tactical scheduling questions.

Problem Description

The physical rail network of yards and tracks defines the allowable routes for train movements. The train network that overlays the rail network defines how traffic can move from yard to yard. Ideally, the operating plan simultaneously allocates physical rail network resources to trains and allocates scarce train space to traffic flows in a way that minimizes operating costs while meeting customer requirements.

Various researchers have considered the operating-plan problem: Assad [1980, 1982], Crainic, Ferland, and Rousseau [1984], Haghani [1989], and Keaton [1989, 1992]. However, they all consider a single representative day's traffic on small networks and produce daily train frequencies. For Santa Fe, the OPM produces a weekly timetable for a widely fluctuating daily arrival pattern and a considerably larger rail network.

Problem Decomposition

The operating-plan problem can be decomposed into two parts, the train-timetable problem and the traffic-assignment problem, which determines traffic flows given a train timetable.

The train timetable, which defines the

time-space network over which traffic may flow, is made up of train schedules and routes. The schedule indicates the days and times a train runs a route. The route defines the origination, termination, intermediate stops (work), speed, and expected travel and yard times of a train. The maximum tonnage of the train and the horsepower (locomotives) assigned to pull it determine train speed and travel time. The train does scheduled work at intermediate locations along its route when it picks up or sets out traffic.

Given a fixed train timetable, the traffic-assignment problem maps an itinerary, or series of trains, for each load from its origin to its destination between its origination time and required time at destination. Each load constitutes a unique flow through the network, creating a large-scale multicommodity flow problem.

Operating-Plan Model

The OPM minimizes the schedule-related costs of service subject to meeting goals for customer service, and engineering and physical plant constraints. (The mathematical formulation is in the appendix.)

The physical capabilities of rail terminals and rail lines constrain the number of trains that may run in the timetable. Terminals and lines have maximum capacities for the number of trains they can accommodate in a time period. These maximums avoid the congestion that causes a railway to fail to meet its timetable. The timetable problem incorporates physical capacity constraints of yard and line.

The traffic-assignment subproblem incorporates engineering constraints on the maximum traffic per train. We set constraints on train size so that trains are not delayed by the amount of traffic on the train. By avoiding yard and line congestion and oversized trains, a railway can plan achievable train-running times independent of traffic patterns. However, a feasible timetable does not imply it moves traffic according to customers' requirements.

Customer-service constraints dictate the time each load in the system must arrive at its destination. Santa Fe has worked with its customers to understand and formalize their service goals and has stratified the customer base accordingly into appropriate service levels so that it can offer services that best suit its customers. This stratification improves Santa Fe's ability to allocate scarce train space in such a way that it can minimize its failures to serve its customers.

Service design in freight rail has been myopic.

The OPM minimizes the costs of crew, horsepower, fuel, handling (or blocking), and equipment and locomotive time. Each train has a fixed cost for crew that is a function of only the distance a train travels. Because crew cost is independent of the number of loads a train carries, it can be dealt with in the train-timetable problem prior to the traffic-assignment problem.

The marginal cost per load, assessed in the traffic-assignment problem, is based on the additional fuel and horsepower requirements the train needs for each additional load. The fuel and horsepower for each load depend on the speed of the train; high-speed trains have a higher marginal cost per unit than low-speed trains. In addition, the high-speed trains have lower capacity and the same fixed crew cost; thus their minimum average total cost per unit is higher. The lower marginal cost and higher capacity give the railway strong cost incentives to use the slower trains, but this is not always feasible when the customer requires fast service.

The costs for equipment and locomotive time reflect the railway's desire to move the traffic across the network quickly so that it can use the equipment sooner for other loads. These costs are a function of the time each load or train spends in transit and in intermediate yards between its origination and its termination. Although the costs for equipment time are a significant percentage of total scheduling costs, the additional costs for equipment time and locomotive time never outweigh the savings in fuel and horsepower in using slower trains. However, if all other things are equal, it is better to move traffic sooner rather than later.

Yards assess handling or blocking costs to cover the additional complications and resources required for trains hauling multiorigin and multidestination traffic. A block is a unique origin-destination pair of traffic on a train. As the number of blocks on a train grows, so does the complexity of its handling. Handling costs give railways incentive to create trains that carry homogeneous traffic; however, by mixing traffic they can create fewer and longer trains, which saves crew costs.

The Problem Size

The OPM produces as output a full weekly timetable of trains. The model represents each hour of the week for every possible train as a binary decision variable, where 1 indicates a train and 0 indicates no train. The size of the problem is directly affected by the number of possible trains in a schedule, which for Santa Fe's intermodal network is approximately 8,040,000 possible trains [Gorman 1998]. Given over 8,040,000 possible trains, there are $2^{8,040,000}$ possible train combinations, corresponding to over two million zeros in the number of possible train timetables.

To reduce the problem size for Santa Fe, the Santa Fe service design team specified a menu of allowable train routes from which the model may choose, similar to those of Keaton [1992] and Crainic, Ferland, and Rousseau [1984]. Although the menu of train routes constrains the solution space, it provides three important advantages: a speedy solution, simplified solution technique, and improved applicability. First, the menu allows us to reduce the problem size from the more than eight million possible train routes to less than 200 likely possibilities (which is reasonable, given that currently in Intermodal there are less than 100). Although the problem still has $2^{33,600}$ possible timetables, the train-route menu reduces the number of possible timetables by 99.5 percent. Second, because train routes determine work locations before traffic is assigned, assigning traffic does not change train stops and travel times; thus, the train-route menu allows us to separate the train-scheduling from the traffic-assignment problem. Finally, the railway can obtain more practical solutions by allowing some user constraints through the train-route menu. For example, such constraints as "Never start a train in Fort Madison" and "No east-bound trains may pick up traffic in Kansas City" are easily implemented through the train-route menu.

The Solution Technique

An iterative, two-step solution technique solves the operating-plan problem. Step one generates a train timetable. Step two assigns traffic given the fixed schedule. The algorithm then returns to step one, which uses the information obtained from the sum of the previous train and traffic-flow costs and generates a new train schedule. The search concludes after it finds no improvement for some number of iterations or reaches some specified maximum number of iterations.

The Train-Timetable Problem

Two search techniques combine to solve the train-timetable problem: genetic search and tabu search. Both techniques have proven effective at achieving successively better solutions while avoiding local optima. For this problem, both approaches were ineffective separately but performed well in tandem.

It is impossible to solve the traffic-assignment subproblem in an efficient way.

Genetic search [Goldberg 1989] borrows from Darwin's theory of survival of the fittest. From a population of candidate train schedules, it is most likely to choose the healthiest (in terms of objective function value) to cross over with other schedules, carrying the best characteristics of the schedules on to the next generation. It relies on random selection for mating and random mutation of schedule characteristics to avoid local optima. Unfortunately, as in evolution, this approach can be painfully slow at producing useful populations of train schedules for this large problem.

Tabu search constructs [Glover 1989, 1990] enhance the performance of genetic search. Tabu search uses information from previous iterations to search in a positive direction. However, a tabu list prohibits the algorithm from undoing recent changes to the schedule or revisiting recent solutions to avoid cycling and local optima. Aspiration criteria may override the tabu if it is clear that the change is in a positive direction. Because it is difficult to identify a good move in the operating-plan problem, tabu search did not do well on its own.

In conjunction with genetics, tabu search acted as a guided mutation. For each genetically produced schedule, the algorithm produces a clone and then alters it using tabu-search techniques. If the tabu change is unsuccessful, the model most likely discards the clone in the next genetic population; if it is successful, tabu search dramatically improves the speed of genetic search [Gorman 1998].

The Traffic-Assignment Problem

It is impossible to solve the traffic-assignment subproblem in an efficient way, and the OPM cannot use iterative solutions from the literature because they are too time consuming. Further, Santa Fe uses a priority-based scheme for assigning traffic to trains in real time. Thus, the OPM uses a priority-based, shortest-path heuristic to assign traffic to trains; it is time efficient, performs well, and most closely reflects railroad-operating reality. Using this heuristic, the OPM assigns traf-

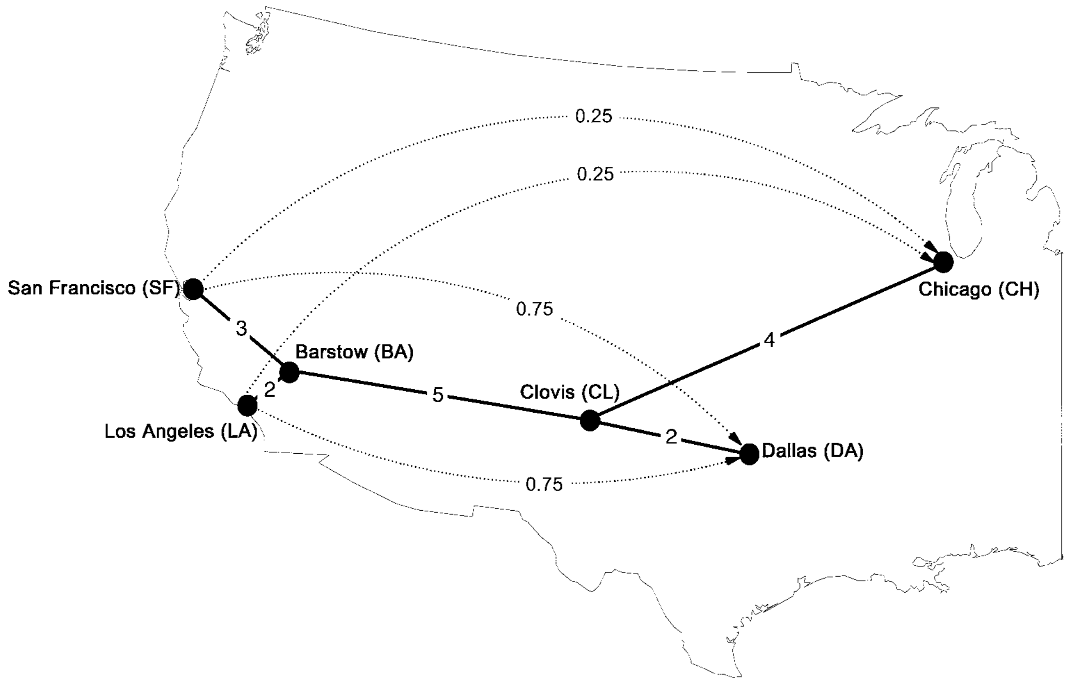

Figure 3: A scaled-down illustrative problem shows the number of trains' worth of traffic per day on dotted arcs and the number of crews per track segment in solid lines. Table 1 shows four possible train schedules for this network.

fic to least-cost trains prioritized by a shortest-goal-first rule. As some trains reach their load capacity, the OPM finds second-choice itineraries for the lower-priority and longer-goal traffic.

An Illustrative Example

We can illustrate the problem faced by the OPM through a small six-location example (Figure 3). To keep the example manageable, we will assume the railway provides daily service for all traffic and all itineraries meet customer requirements. Further, assume all trains run at the same speed, so fuel and locomotive requirements do not affect costs. In this example, we may run zero, one, or two trains per day. Let the objective be to minimize the sum of crews, handling, and total blocks required to move all traffic each day.

Crew costs are a function of the track segments traversed. Figure 3 shows crews by segment; Table 1 shows total crew requirements for each train. Handling is a switch between trains at an intermediate yard; it reflects the additional cost for equipment time required between trains. A handling cost is assessed for all trains terminating at an intermediate yard. Blocks are the number of unique origin-destination pairs of traffic on each train and reflect the additional yard resources required to manage multiblock trains. Only one block exists in direct train service, and in this problem, two blocks are created for any train terminating or originating at an intermediate yard because of

Train route menu	Train route costs: Crews	Blocks	Handling	Sample train schedules: #1	#2	#3	#4
SFBA	3	2	1	0	1	0	1
SFCL	8	2	1	0	0	1	0
SFCH	12	1	0	1	0	0	0
SFDA	10	1	0	1	0	0	1
LABA	2	2	1	0	1	0	1
LACL	7	2	1	0	0	1	0
LADA	9	1	0	1	0	0	1
LACH	11	1	0	1	0	0	0
BACH	9	2	0	0	1	0	1
BADA	7	2	0	0	2	0	0
CLCH	4	2	0	0	0	1	0
CLDA	2	2	0	0	0	2	0
BACL (N/A)				0	0	0	0
Total crews				42	28	23	33
Total blocks				4	10	10	8
Total intermediate handling				0	2	2	2
Total cost (C + B + H)				**46**	**40**	**35**	**43**

Schedule #1: Direct service for all traffic
Schedule #2: Handle all traffic at Barstow
Schedule #3: Handle all traffic at Clovis
Schedule #4: Direct service for Dallas, handle Chicago traffic at Barstow

Table 1: Four of the many possible schedules for the scaled-down network of Figure 3 are shown (1 = train; 0 = no train). The simplest schedule, Schedule 1, runs direct, single-block trains (LACH, LADA, SFCH, SFDA) but ignores opportunities for combining traffic from multiple trains into one at Barstow and Clovis. Schedule 2 shows the possibility of moving all traffic out of San Francisco and Los Angeles to Barstow (SFBA, LABA), where it is combined into trains to Chicago and Dallas (BACH, BADA). Schedule 3 shows the same flows to Clovis (SFCL, LACL), where it is split into Chicago and Dallas trains (CLCH, CLDA). Schedule 4 serves Dallas in direct train service (SFDA, LADA) and combines Chicago traffic at Barstow (SFBA, LABA, BACH).

the combination of traffic of two origins or destinations.

Consider a representative day's eastbound traffic volume (given in trainloads in Figure 3) from Los Angeles and San Francisco to Chicago and Dallas. This example has 13 possible trains (Table 1). The service-design expert may, for example, reduce the options by excluding the BACL train due to some constraint, such as locomotive availability, which is otherwise difficult to represent to the model. The OPM takes the remaining 12 train routes in the menu of trains and finds a least-cost timetable of trains that carries the traffic.

The trade-offs are between the larger number of crews required to move the traffic in single-train service versus the increased blocking and handling from combining traffic at intermediate locations in multitrain moves. Table 1 illustrates the scheduling trade-offs.

In the four options listed in Table 1, the OPM shows it is best to consolidate traffic at Clovis (Schedule 3). In this schedule, the trains traveling the greatest distance are fullest, creating the greatest return for the crews spent. The best solution depends heavily on the volume between origins and destinations. For example, if all traffic flows were integer train volumes, direct service would have been the least-cost choice. Alternatively, if the LACH volume was .75 and SFDA volume was .25, then the handling would better take place in Barstow because full trains run from Barstow to destination.

The introduction of traffic-originating intermediate terminals, time of arrival and service requirements of customers, and day-of-week fluctuations in traffic vastly complicate the problem. This simple example demonstrates the complicated network-wide interaction of train schedules and traffic flows.

Applications of the OPM

We designed the OPM to take a clean-slate view of service design. When we applied the OPM to the entire intermodal-service plan while restricting allowable train routes to only those currently in use at Santa Fe, we found the potential for a four-percent reduction in schedule-related costs coupled with a six-percent reduction in late service compared to current planned services. Given the size of schedule-related operating costs at Santa Fe Intermodal, these percentages represent significant dollar savings. More important, the simultaneous reduction in costs and late service indicates a major breakthrough in the cost-service trade-off experienced with incremental modifications to the service plan.

However, because the model produces operating plans much different from current Santa Fe operations, we found it difficult to gain managers' full confidence in its feasibility. Further, because of the interdependencies of the trains in the timetable, instituting a subset of model recommendations can be hazardous. Finally, it is a matter of practical impossibility to make sweeping, wholesale changes to the service plan without destroying productivity in daily operations. To develop managers' confidence in the model results and to make the model more tactically useful, we modified it and applied it to smaller-scale problems in operating-plan development.

We limited the model to cover trains in particular corridors of business to analyze key subsets of the Santa Fe network. At the suggestion of service-design experts, we made it possible to hold a portion of the train timetable constant and search for additional trains to support these "untouchable" trains. These limitations on the scope of the model allow Santa Fe to address finely tuned service-design questions.

For example, in our most heavily traveled corridor from Chicago to Los Angeles, our premium train network is well established and not a likely candidate for change because of contractual agreements. We used the OPM to develop secondary train service in the corridor, given a fixed premium train network and traffic flows. The results showed that Santa Fe needed to shift the secondary train network offerings by reducing trailer-train service and providing more stack-train (containerized) service, thus reducing the total number of trains required to serve our customers.

The model has helped open people's eyes to new scheduling paradigms. Typically, a train route runs at the same time of day every day of the week it runs. In our Birmingham-to-Los Angeles market, the OPM indicated that a five-day-per-week train could be reduced to only four runs per week by running a train every 36 hours. By running the train Monday and Thursday at 6 AM, and Tuesday and Friday at 6 PM instead of Monday through Friday at 6 PM, we could serve all five days of traffic on time and annul the Wednesday train start. Only the established standard of scheduling in 24-hour cycles drove the need for a fifth train.

Given the success of the model on more tactical problems, we applied it to more strategic operating questions, not so much as a turnkey schedule developer, but as a means to test the cost and service effectiveness of different operating philosophies. For example, we wanted to quantify the savings of going to a hub-and-spoke operation. We estimated that Santa Fe would make considerable savings in crews by running larger trains in and out of a hub in the center of the railroad. While the OPM showed that the crew savings would be significant, it found those savings to be more than offset by additional yard-handling and locomotive- and equipment-time costs incurred under the hub-and-spoke scenario. The model indicated that 80 percent of our intermodal traffic on major corridors should be handled in direct, one-train service.

We have also tested other operating strategies. We developed a shuttle-train-service schedule between Chicago and Kansas City with the help of the OPM. We then tested the shuttle-service philosophy across the railroad, with short-route, high-capacity trains running at higher frequencies and uniform, lower speeds. Each train could carry many blocks of traffic which were sorted at intermediate stations. We found train operations could be simplified and service improved with more frequent, multidestination departures making up for low-speed trains.

We have also used the OPM to develop schedules around problems with the physical track and yard network. For example, maintenance of way (track repair) operations in single-track areas can shut down or reduce service during certain time windows. When a major project began on the Tehachepi tunnel on Southern Pacific lines in Northern California (over which Santa Fe trains travel), we used the OPM to develop train service around the maintenance windows. The solution showed the trade-off between the later departure of high-speed trains (which traversed the Tehachepi tunnel after the maintenance window), which have high cost and high customer-service levels, and earlier departure of low-speed trains (before the Tehachepi window), resulting in lower costs but higher customer failure some days of the week. Similarly, we have used the OPM to find scheduling alternatives to trains doing work in Argentine yard in Kansas City because of the congestion delays they experienced there because of reconstruction of the yard. Because Ar-

The model has helped open people's eyes to new scheduling paradigms.

gentine yard was not available for handling traffic, the OPM showed we needed additional trains to support the traffic.

Conclusion

Santa Fe Intermodal has used the operating-plan model to study many major changes in rail operations: to predict train volumes based on long-term forecasts, to quantify the impact of containerization of intermodal business on train operations, and to develop a cost basis in contract negotiations for large amounts of incremental business. It is most useful when expert opinion and historical operations provide insufficient information for indicating impacts of major changes on operations. It is a flexible tool that has been used to address philosophical, strategic, and tactical problems in the railroad's service design.

The operating-plan model is designed for developing and evaluating any radical departure from current operations. For example, it could be used to develop service plans in times of strike, flood, or merger. As a clean-slate approach, it does not rely on incremental adjustments to existing operating plans to arrive at improved service designs. The operating-plan model provides Santa Fe with the insights it needs to improve customer service while reducing costs.

APPENDIX: Mathematical Formulation: The Operating Plan Problem

Minimize $\Sigma_t \; FC_t T_t \; + \; \Sigma_s \Sigma_t \Sigma_{l \epsilon R^s} \; MC_{tl} D^s_{tl}$

Subject to

$D^s_{tl} > 0$ for all s, for $l \in R^s$.

$D^s_{tl} \in \delta$ for all $s, t, l \in R^s$.

$C^*_t T_t \geqq {}_s D^s_{tl}$ for all $t, l \in R^s$.

$T_t \in \Omega$ for all t.

$T_t = (0,1)$ for all t.

$D^s_{tl} = (0,1)$ for all $s, t, l \in R^s$.

Indices:

t: index for trains.
s: index for demands.
n: index for nodes (yards on system).
l: index for links (track between consecutive yards).

Sets:

R^s: the set of links over which load s travels (itinerary).
S_t: the set of demands carried by train t.
δ: the set of service restrictions for demands.
Ω: the set of congestion restrictions for trains.

Decision Variables:

T_t: binary variable, 1 if providing T_t, 0 otherwise.
D^s_{tl}: binary variable, demand to be serviced: 1 if on train t over link $l \in R^s$, 0 otherwise.

Parameters:

FC_t: fixed cost of providing train t (crew cost).
MC_{tl}: marginal cost of carrying demand on train t, over link $l \in R^s$ (includes fuel, handling, locomotive, and equipment costs).
C_t: capacity of train t, in units of demand.

References

Assad, A. A. 1980, "Modelling of rail networks: Toward a routing/makeup model," *Transportation Research,* Vol. 14B, No. 1, pp. 101–114.

Assad, A. A. 1982, "A class of train-scheduling problems," *Transportation Science,* Vol. 16, No. 3 (August), pp. 281–310.

Crainic, T.; Ferland, J.; and Rousseau, J. 1984, "A tactical planning model for rail freight transportation," *Transportation Science,* Vol. 18, No. 2, pp. 165–184.

Glover, F. 1989, "Tabu search—Part I," *Journal on Computing,* Vol. 1, No. 3, pp. 190–206.

Glover, F. 1990, "Tabu search—Part II," *Journal on Computing,* Vol. 2, No. 1, pp. 4–32.

Goldberg, D. 1989, *Genetic Algorithms in Search, Optimization and Machine Learning,* Addison-Wesley, Reading, Massachusetts.

Gorman, M. 1998, "An application of genetic and tabu searches to the freight railroad operating plan problem," *Annals of Operations Research*, Vol. 18, 51–69.

Haghani, A. E. 1989, "Formulation and solution of combined train routing and makeup, and empty car distribution model," *Transportation Research*, Vol. 23B, No. 6, pp. 433–452.

Keaton, M. H. 1989, "Designing optimal railroad operating plans: Lagrangian relaxation and heuristic approaches," *Transportation Research,* Vol. 23B, No. 6, pp. 415–431.

Keaton, M. H. 1992, "Designing railroad operating plans: A dual adjustment method for implementing Lagrangian relaxation," *Transportation Science,* Vol. 26, No. 4, pp. 263–279.

R. Mark Schmidt, AVP-Strategic Studies, The Atchison, Topeka and Santa Fe Railway Company, PO Box 1738, Topeka, Kansas 66601-1738, writes: "The Operating Plan Model (OPM), developed by Mike Gorman, has proven to provide an accurate, reliable alternative to the traditional service design process for our intermodal traffic base. As a result of the OPM, these traditional methods and outcomes have been challenged and a set of train schedules which has improved the quality of our service offerings at reduced cost levels has resulted.

"Specifically, the OPM has been used to analyze intermodal scheduling in the following areas: Southern California; Northern California; Clovis, New Mexico, Hubbing Operations; Service to the Southeast (Memphis and Birmingham); East Saint Louis; and Kansas City.

"Obviously, as with any major deviation from traditional processes, the acceptance of the OPM has been a gradual one. Recent successes of the model are building confidences and as a result, the OPM is being interwoven into the intermodal service design process at Santa Fe."