DESIGN AND IMPLEMENTATION OF AN INTERACTIVE OPTIMIZATION SYSTEM FOR NETWORK DESIGN IN THE MOTOR CARRIER INDUSTRY

WARREN B. POWELL

Princeton University, Princeton, New Jersey

YOSEF SHEFFI

Massachusetts Institute of Technology, Cambridge, Massachusetts

(Received May 1986; revisions received July 1987, June 1988; accepted September 1988)

We describe the design and implementation of an interactive optimization system for routing freight over a less-than-truckload motor carrier network. We formulate a very large, mixed integer programming problem, and develop a decomposition strategy based partly on the mathematical structure of the problem as well as a range of important, real-world issues and constraints. Then we develop and implement a local improvement heuristic in such a way as to keep the "man-in-the-loop," using the analyst to make judgments regarding certain complex constraints and tradeoffs. Important aspects of the system include a range of modeling approximations to keep the problem tractable and the way the analyst evaluates the quality of the different numbers. The package was implemented and is currently being used on an ongoing basis by a major motor carrier. An overview of the major elements of the package is given as well as a summary of important implementation issues that arose during the three year project.

The load planning problem of less-than-truckload (LTL) motor carriers consists of determining how to consolidate flows of small shipments over a network of *breakbulk* terminals to minimize transportation and handling costs while maintaining level of service. In a joint project, funded by IU International Inc. and carried out at Princeton University and the Massachusetts Institute of Technology, a computer package was developed for solving the load planning problem and implemented at a subsidiary of IU, PIE Nationwide (previously Ryder/PIE Nationwide, and originally, Ryder Truck Lines). PIE Nationwide is a national LTL motor carrier that operates with over 300 terminals throughout the United States. The development of this package, dubbed APOLLO (Advanced Planner of LTL Operations), required exploiting the mathematical structure of the problem in order to develop a process that met the planning needs of the carrier. The implementation of APOLLO, however, required a careful merging of the computer's ability to solve quickly certain very large, but mathematically well behaved, problems with the analyst's skills in recognizing spatial patterns as well as making difficult to quantify tradeoffs.

This paper exposes the broad range of objectives and decision variables that need to be balanced to develop an implementable load plan. Critical to the success of the project is the application of classical optimization techniques to certain well chosen parts of the problem. This factor distinguishes APOLLO from other interactive tools which simulate a user generated solution. Equally important were a series of compromises that recognize the inability and the lack of desire to incorporate many extremely complex and poorly defined issues into the model. Instead, tradeoffs involving these issues were left to the judgment of an analyst with detailed knowledge of the carrier's operations. The combination of basic optimization and interactive involvement is applicable to the solution of many complex problems in the real world.

The load planning problem has been the subject of a considerable amount of attention. The approach originally used at Ryder Truck Lines is based on a simple network building heuristic that links to the network if there is sufficient flow, calculated using a predetermined set of rules. A similar method was used by Multisystems (1979), and Temple, Barker and Sloane (1978). Barker, Sharon and Sen (1981) present

Subject classification: Networks/graphs: interactive local improvement for network design. Transportation: network planning for motor carriers.

0030-364X/89/3701-0012 $01.25

a model that uses simulation in conjunction with linear programming in an iterative process to solve the load planning problem. Starting with a feasible load plan, the model iteratively analyses all lanes that are served directly and prescribes where equivalent or better service could be provided at less cost via breakbulk service. Also, there is a reverse process for lanes scheduled for breakbulk service; the model indicates where direct service is cheaper and better. This provides the basis for subsequent iterations. Flows over the network are then balanced in an optimum least-cost manner using a linear program. The model was used by ANR Freight in a strategic planning study of breakbulk requirements and for evaluating different network configurations, and continues to be used for load planning.

More recently, Powell and Sheffi (1983) outline a local improvement heuristic that serves as a precursor to the approach described in this paper. Roy (1984) uses an objective function that combines transportation costs and level of service constraints, where the revenues in a traffic lane are effectively assumed to be lost if travel times in that lane exceed the service time constraint. Simple queueing models are used to estimate delays at terminals as a function of frequency. The method seems to work well on very small networks (30 terminals). Balakrishnan (1984), and Balakrishnan and Graves (1985) determine a rigorous cost lower bound for LTL freight flows over a network from the perspective of a shipper facing declining marginal costs as a function of flow. The approach to the load planning problem in this research is similar in structure to the multicommodity fixed charge network problem as studied by Magnanti and Wong (1984), and Lamar, Sheffi and Powell (1984). Lamar and Sheffi (1985) extend the previous work by considering link cost functions that are piecewise convex in addition to the fixed charge. Such an approach can be used to model cost functions with a flat plus a linear shape, first used by Powell and Sheffi (1983) for this problem. Finally, Powell (1986) presents a local improvement heuristic that combines the general flavor of the search algorithm in Powell and Sheffi (1983) and the work on solving the routing subproblem in Powell and Koskosidis (1984, 1986).

While much of this work served the important purpose of investigating specific technical issues in some depth, no summary has been made of how the actual implementation affects the structure of the algorithms and the architecture of the software. Also unreported is how the work is currently being used and its effect on planning at PIE. A detailed reading of previous papers written by the authors on the subject leaves the impression that APOLLO is a classical *black box* model, an approach that would have failed miserably in implementation. The objective here is to synthesize the algorithms and their implementation and to expose the important interaction between algorithmic approach, software architecture and the implementation of the package. A number of specific issues are covered which are common to any large, complex model implementation. These include the following.

- The role of interactive optimization which involves a true human in the loop. This process allows the model to make certain approximations in both the cost model and the constraint set which would be corrected by a knowledgeable user during the optimization process.
- The choice of appropriate levels of accuracy and complexity. It was important that the model be sufficiently realistic that the large majority of recommendations be accepted by management. At the same time, it was necessary to communicate the logic behind the calculations to nontechnical management in such a way that the numerical calculations can be verified manually.
- The distinction between hard and soft numbers. Certain cost calculations and constraints were more accurate than others (or more easily measurable, in the case of the constraints) as a result of details that could not be incorporated in the model (either due to lack of data or computational constraints). The design of the package requires the analyst to be in a position to make these evaluations.
- The use of decomposition. The natural structure of the problem, as usual in transportation, lent itself to a hierarchical solution approach that simplified the algorithms, as well as significantly accelerated their speed. In addition, this decomposition was useful in separating the more accurate cost calculations from the less accurate ones.

To understand the problem and some of its complex constraints, Section 1 gives a brief review of LTL network operations. Section 2 casts the problem in a classical optimization framework, producing a large, mixed integer programming problem with an unique structure. Section 3 argues for a particular decomposition strategy based as much on several real-world concerns as on mathematical arguments. Section 4 briefly outlines how to solve the subproblems, and Section 5 presents the overall solution approach that encompasses both the interactive elements and the search algorithms. Section 6 provides an overview of the architecture of the software, which plays a significant role in the development of an interactive optimization process. Section 7 summarizes some of the broader implementation issues that play an important role in the successful adoption of this new methodology. Finally, Section 8 reviews the actual impacts the model has had on the organization.

1. Background Information

The load planning problem at PIE is a tactical planning exercise to determine the routing of shipments through a specified set of consolidation (*breakbulk*) terminals. The result of the load plan is a determination of how to route trucks and a freight movement plan (FMP), which specifies how to route individual shipments. The FMP is a set of instructions of the form, "Shipments at terminal *i* headed for destination *d* must be put on a trailer headed for breakbulk *j*."

As a rule, trailers are loaded at one terminal and completely unloaded at the destination. If the destination terminal for a trailer is not the final destination for a shipment on that trailer, the shipment must be sorted and reloaded onto another outbound trailer. It is the load plan that specifies on which outbound trailer the shipment is placed. Note that the load plan applies only to the *less-than-truckload* freight (comprising shipments from 500 to 10,000 pounds), which must be consolidated on the trailers. Most LTL carriers also haul a significant amount of *truckload* freight, where an individual shipment requires an entire trailer. Since truckload freight does not need to be consolidated, it is handled separately.

To properly describe the load planning problem, it is useful to understand some elements of LTL operations. An LTL network consists of end-of-line terminals, where most of the freight originates and terminates, and breakbulks, which handle the unloading, sorting and reloading of the freight from one trailer to the next. Reflecting restrictions on driving time, the movement of tractors between terminals must follow the *line operations network*; Figure 1 shows an example. The length of any link in this network cannot exceed what a driver can cover in 10 hours driving time. The nodes of the line operations network, consisting of all the terminals as well as additional relay points, are points at which drivers are often changed.

When a trailer is loaded at terminal *i* and unloaded at terminal *j*, the carrier is said to offer *direct service* from *i* to *j*. Offering direct service between two terminals generally implies a regular operation with trail-

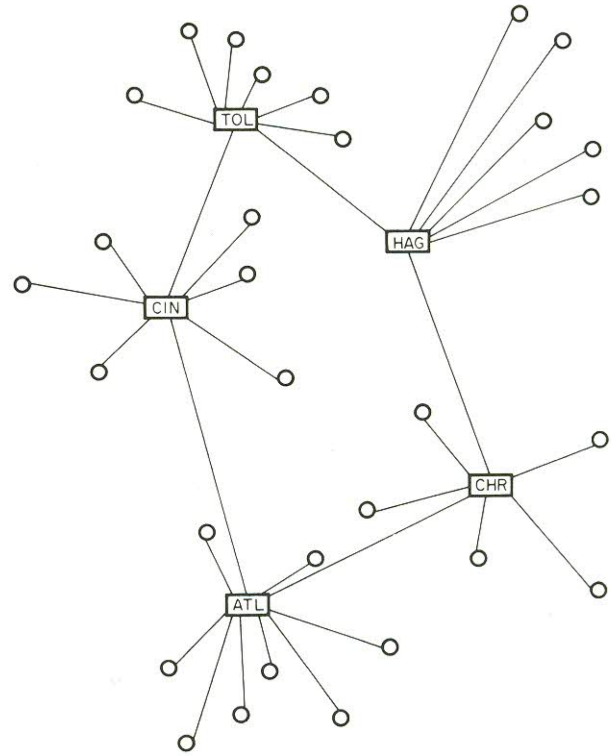

Figure 1. Illustrative line operations network.

ers leaving at least two or three times per week. At lower frequencies, some shipments incur unacceptably long delays waiting for the trailer to leave.

A major component of the load planning problem is determining to which pairs of terminals the carrier should offer direct service. Since a single direct service between two terminals may comprise several links of the line operations network, it is useful to follow Powell and Sheffi (1983) and introduce the *load planning network*; Figure 2 illustrates an example. In principle, the load planning network might consist of links that connect every terminal to every other terminal. For large national networks, however, direct end-of-line to end-of-line movements are rare and, for discussion purposes, can be ignored. Thus, the set of load planning links can be viewed as all links that begin or end at a breakbulk. Of this total remaining set of load planning links, only about 10 to 20% will be used for direct service. The problem is to determine which ones to use.

The day-to-day operations of LTL networks encourage the development of certain relationships between the terminals. Each end-of-line is typically associated with one *primary break*, which is usually the closest breakbulk, depending on the directionality of the freight. The end-of-lines served by a given primary

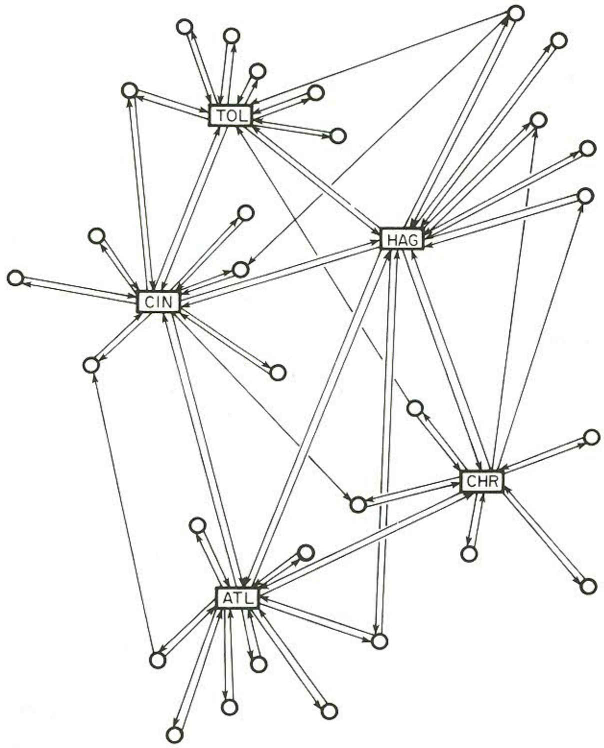

Figure 2. Illustrative load planning network.

breakbulk are referred to as the satellites of that breakbulk. Direct service is offered virtually always between a satellite and its primary break, with a frequency of at least one departure per day. This policy maintains regular service into and out of the end-of-line and simplifies daily vehicle and manpower routing problems. Trailers outbound from an end-of-line destined to more distant breakbulks often relay at the primary breakbulk to change drivers.

Perhaps the most important tradeoff in the design of a load planning network is between costs and level of service. Carriers generally manage level of service by requiring that the frequency of service over each load planning link satisfies a specified minimum. In most cases, direct service will not be offered between two terminals unless it is possible to fill at least the minimum number of trailers per week, with the exception of movements between satellites and primary breaks where departures occur regardless of the amount of freight on the trailer. Typical weekly minimums for breakbulk to breakbulk moves or from an end-of-line and nonprimary break are 3–5 trailers per week. The use of minimum frequencies might seem like a heuristic approach for handling level of service constraints, but it is, in fact, very realistic in terms of how many carriers actually operate. For this reason, level of service is more an output than an input of the process.

As a result of the minimum frequencies, it is possible to write the flow of trailers, F, as a function of the flow of trailerloads of freight, x, as

$$F_{ij}(x_{ij}) = \begin{cases} \max(M_{ij}, x_{ij}) & \text{if } x_{ij} > 0 \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

where

x_{ij} = the trailerloads of freight per week from i to j,

F_{ij} = the trailers dispatched per week from i to j.

M_{ij} = the minimum frequency if direct service is offered from i to j.

Equation 1 has the shape given in Figure 3, which implies a fixed charge of M_{ij} if direct service is offered from i to j. Once this fixed charge is incurred, the marginal cost is zero until $x_{ij} = M_{ij}$, at which point the number of trailers increases linearly with the flow of freight (measured in trailers). One might assume that the function should increase in discrete increments once the flow exceeds the minimum. In reality, however, the dispatching function becomes a pure go-when-filled operation when flows exceed the minimum, and all flows must then be viewed as weekly averages. Thus, if $F_{ij} = x_{ij} = 3.6$ then, on average, the carrier will send 3.6 trailers per week.

2. The Optimization Formulation

Before describing the optimization formulation in mathematical terms, it is necessary to review the

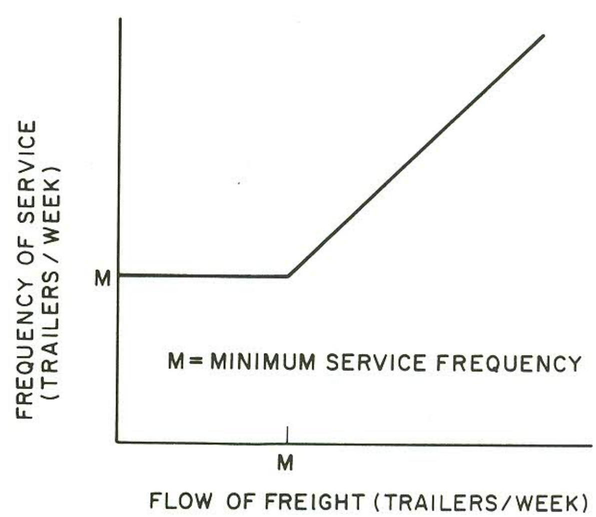

Figure 3. Departure frequency as a function of flow showing the minimum load constraint.

decision variables, costs and constraints. The decision variables can be grouped into three different categories:

Primary — where to offer direct service,
Secondary — how to route the LTL freight over a given network of direct services (the load planning network), and
Tertiary — how to route the truckload freight and balance the empty trailers over the line operations network.

These decision variables must be chosen to minimize the following costs.

Direct transportation. The costs of pulling loaded trailers over each link of the network.

Handling at breakbulks. The cost of unloading, sorting and reloading freight as it passes through each breakbulk.

Handling at origin. The total amount of freight passing through an end-of-line is fixed, but the unit cost depends on how much sorting is required. If an origin end-of-line loads trailers directly to only one breakbulk, no sorting is required and costs are lower than if trailers are being sent to multiple breakbulks. This unit cost is specified as a step function for loading from the origin to one, two, three or more breakbulks.

Empty balancing. The cost of balancing empty trailers to ensure overall conservation of flow.

Truckload. The cost of moving truckload freight over the network.

These costs relate to the load plan in some way. Other operating costs, such as pickup and delivery or handling at destination terminals, were not included because they are independent of the load plan.

In considering the constraints on LTL operations, it is instructive to first describe those not included in our problem formulation. The major types of constraints *excluded* are the following.

Level of service constraints. For reasons described earlier, no attempt was made to enforce explicit constraints on origin to destination travel times. Aside from making the problem ungainly large, the use of minimum frequencies represents a much more realistic approach to this problem.

Breakbulk capacities. Breakbulks have definite limits on how much freight they can handle, but the limits are not hard constraints, and they depend as much as *when* the freight will arrive at the break as how much freight is passing through it. Since the model uses steady state flows, it is not possible to include reliable breakbulk capacity constraints.

Link flows. No capacity was placed on how much freight could move over any given link. This is equivalent to assuming that there is no fleet size constraint, a realistic assumption in today's overcapacity conditions.

Aside from the usual flow conservation and nonnegativity constraints, there are two important sets of constraints that govern how freight is routed over the load planning network. These are the *tree constraints* and *cluster constraints*. Tree constraints specify that the flow from origin to destination must follow a directed spanning tree into the destination. This not only provides an unique path between every pair of terminals, but one that is independent of the origin of the shipment.

The cluster constraints reflect the relationship between satellites and primary breaks. In a large national network, terminals are grouped around their primary breakbulks. In general, all the freight from one group of terminals (a breakbulk and its satellites) going to a terminal in a particular destination group, and passing through the primary break for that group, should follow the same freight movement plan. There are, however, exceptions to this rule.

The cluster constraints are discussed in depth in Powell and Koskosidis (1986) and can be summarized as follows.

1. Freight from a terminal O (which may be an end-of-line or a breakbulk) into an end-of-line D must follow the same path as freight from O to the primary breakbulk B of D, with one exception: freight may follow a different path if that path does not pass through B.
2. Freight passing through B that is not destined to B or any of its satellites need not follow the routing of freight destined at B. Freight destined to a satellite of B is often viewed as *belonging* to B while freight moving through B to another breakbulk is viewed independently and can follow its own routing.
3. Freight into an end-of-line D with more than one primary break must follow the path that either avoids all primary breaks or it must choose a valid path (as defined by constraint 1) through one of the primary breaks.

The inclusion of these detailed routing constraints played a significant part in the eventual acceptance of the model. This was true for several reasons. First, they allowed us to read in the carrier's current load plan and match it exactly within the computer, regardless of whether it followed the *least cost* path or not. It also allowed us to duplicate unusual routings that were required for real-world reasons, such as international freight that needed to move through a specific port or constraints on driver movements and breakbulk capacities. Second, and almost more important, the ability to enforce detailed routing constraints provided a psychological level of control over what the computer was doing. At all times, the industrial engineers using the model knew they retained ultimate control over the solution. This was particularly important because at the end of an optimization session the computer would automatically update the files for routing shipments, which were then sent directly to the field. At the same time, the additional routing constraints posed some particularly difficult optimization problems that affected the approach of the solution.

The load planning problem can be cast as a conventional, large-scale mathematical programming problem. This is presented in detail in Appendix A1. The discussion here is intended only to provide a sense of the problem's elements and structure. We use the convention that all vectors are column vectors. We define

$y = \{\ldots, y_{ij}, \ldots\}$
= the vector of network design variables, where $y_{ij} = 1$, if the carrier is offering direct service from terminal i to j, and 0 otherwise,

$x^s = \{\ldots, x^s_{ij}, \ldots\}$
= the vector of LTL freight flows from i to j with destination s (we require $x^s_{ij} = 0$ if $y_{ij} = 0$),

$x = \{\ldots, x_{ij} \ldots\}$
= the vector of total LTL freight from i to j to all destinations,

$\hat{x}^s = \{\ldots, \hat{x}^s_{ij}, \ldots\}$
= the vector of truckload freight flows from i to j with destination s,

$\hat{x} = \{\ldots, \hat{x}_{ij}, \ldots\}$
= the vector of total truckload freight from i to j to all destinations,

$z = \{\ldots, z_i, \ldots\}$
= the vector of total LTL freight handled outbound from each terminal (thus, if i is an end-of-line, z_i would be the total outbound freight; if i is a breakbulk, z_i would be the total transferred freight plus freight originating at that breakbulk),

$d = \{\ldots, d_i, \ldots\}$
= the vector of net trailer surpluses and deficits caused by the movement of trailers pulling LTL and truckload freight,

$v = \{\ldots, v_{ij}, \ldots\}$
= the vector of empty trailer flows required to balance the flow of loaded trailers.

Also define

$F = \{\ldots, F_{ij}(x_{ij}), \ldots\}$
= the vector of frequency functions, giving the number of trailers moving over each link as a function of the flow.

In addition to these decision variables, the following costs were required:

$c = \{\ldots, c_{ij}, \ldots\}$
= the vector of transportation costs per trailer over each link,

$h = \{\ldots, h_i, \ldots\}$
= the vector of handling costs (per trailer) for each breakbulk ($h_i = 0$ if i is not a breakbulk),

$g = \{\ldots, g_i(N_i), \ldots\}$
= the vector of cost functions, $g_i(N_i)$, where $N_i = \Sigma_j y_{ij}$ is the number of outbound directs out of origin end-of-line i. This is an increasing function of N_i, and typically includes values for $g_i(1)$, $g_i(2)$, and $g_i(3)$, with $g_i(n) = g_i(3)$ for $n > 3$. Assume $g_i(\cdot) = 0$ if i is a breakbulk. Because N_i depends directly on the vector y (the vector y determines the number of outbound directs out of a terminal), we may write $g_i = g_i(y)$.

$q = \{\ldots, q_i, \ldots\}$
= the total flow originating at terminal i.

Total costs were divided into five components, each represented by a separate function. These are

$G_1(y, x)$ = the total transportation costs for moving loaded LTL trailers,
$= c^T F$.

$G_2(z)$ = the total handling costs at the breakbulks,
$= h^T z$.

$G_3(y)$ = the total handling costs at origin end-of-lines,
$= g(y)^T q$.

$G_4(\hat{x})$ = the total costs for moving truckload freight,
$= c^T \hat{x}$.

$G_5(v)$ = the total costs for repositioning empty trailers,
$= c^T v$.

The objective function can now be presented in general terms. We wish to solve

$$\min_{(y,x,\hat{x},v)} G_1(y, x) + G_2(z) + G_3(y) + g_4(\hat{x}) + G_5(v) \quad (2)$$

subject to the following constraints:

1. integrality of the network design vector y;
2. the requirement that LTL flow may move only over links where direct service is being offered;
3. the flow of all LTL flow into a single destination must follow a tree, and must obey the cluster constraints just described;
4. flow conservation for LTL flow;
5. flow conservation for TL flow;
6. flow conservation for all trailers (loaded LTL, TL, and empty).

For a network with 30 breakbulks and 270 end-of-lines (the size of the PIE network during the project), there are 17,000 possible links in the load planning network. Also, as we demonstrate in the Appendix, the requirement that the LTL flows follow a tree requires formulating integer routing variables. For a problem of this size, there would be 27 million integer routing variables. These numbers obviously suggest that the problem may be difficult to solve optimally by a batch model. More importantly, however, real-world considerations make a global approach unacceptable. The large program outlined above, even if it could be solved, involves many approximations which must necessarily be made to avoid making the problem even larger. A sample of these include:

Static freight flows. Freight flows vary by time of day and day of week, in part due to weekend effects. The use of static flows especially compromises the accuracy of the flows of empties, but also makes it difficult to enforce breakbulk capacities and level of service constraints.

Driver constraints. Work rules govern where and when drivers can run. In some cases, drivers will run empty to avoid incurring overnight costs.

Terminal capacities. As mentioned earlier, breakbulk capacities are not represented explicitly. These capacities are difficult to measure, as they depend on when freight moves through a terminal, as well as on details unique to each breakbulk.

Marketing and level of service. The value of maintaining high service levels depends on the individual markets (origin-destination pairs) served.

Given these limitations, the two central modeling issues to emerge during the development and implementation of the solution approach are that

- the solution approach must be able to handle the limitations of the formulation, and
- recommendations made by the model must be believable and verifiable.

The first issue implies that the solution approach must be sensitive to the fact that certain modeling approximations will introduce unavoidable errors. For example, some numbers will be more accurate than others. As a result, the analyst must be in a position to determine how much of the savings from a particular recommendation are based on *hard* (that is, very accurate) numbers as opposed to *soft* ones. In addition, we cannot always accurately measure the degree to which we may be violating certain constraints (such as the level of service or breakbulk capacity), and the modeling system must allow the user to introduce his or her own judgment.

The second issue implies that tradeoffs should not be unnecessarily complex. Attempts to make the model mathematically more precise can frequently make it impossible to understand why it is making certain suggestions (this, of course, depends on the nature of the algorithm and the manner in which information is displayed to the user).

The presence of these approximations has a direct effect on how the problem should be solved. For example, the routing of the LTL shipments over the network of direct services, given by the flows $\{x\}$, is coupled with the routing of the empties. This raises the possibility that the model would choose an alternative routing simply to reduce the cost of routing empties. While mathematically valid, it puts too much weight on an unreliable number (the flow of empties). In addition, it greatly complicates the process of explaining why the model chooses a particular routing. These considerations motivate a solution based on the original global optimization formulation that is modified in several important ways to reflect real-world considerations.

3. The Decomposition Strategy

The global optimization formulation should be augmented by an additional set of constraints that reflects actions not allowed due to the limitations described earlier. Since these constraints are generally difficult to quantify, changes to the load plan that reduce costs must be presented to the analyst for a feasibility check. However, only *major* changes to the network should be reviewed; the possibly thousands of minor changes that might accompany a single major change should be executed reliably, with confidence and require only cursory review.

These issues suggest that the problem be decomposed hierarchically into a network design problem, governing which links are added to or dropped from the network, and a series of subproblems determining the routing of LTL shipments, the routing of TL shipments and the routing of empty trailers to balance the network. Figure 4 illustrates the different decisions and the degree of user interaction associated with each. The values of the network design variables (adding or dropping direct services) are decided one at a time, a process equivalent to a local improvement heuristic. The evaluation of each change in the load planning network requires a reoptimization of the different subproblems to determine the corresponding system impact.

This solution approach can be expressed in terms of a set of optimization problems. The discussion below is again somewhat general, with a more detailed presentation given in Appendix A2. The important concept here is the manner in which a large optimization problem is structured hierarchically as a sequence of optimization subproblems which take advantage of the natural structure of the problem.

Network Design Problem (NDP)

Decision variables (y)
Objective function

$$\min_{\{y\}} G_1(y, x^*(y)) + G_2(z^*(y)) + G_3(y) + G_4(\hat{x}^*) + G_5(v^*(x^*, \hat{x}^*)) \quad (3)$$

Subject to constraint 1.

In problem **NDP**

$x^*(y)$ = the optimal allocation of LTL flows over a given set of directs y,

$\hat{x}^*$ = the optimal flow of TL trailers,

$z^*(y) = z(x^*(y))$ = flows handled at each breakbulk as a function of the vector of LTL flows, x^*,

$v^*(x^*, \hat{x}^*)$ = the vector of optimal empty flows given the optimal flow of loaded LTL and TL trailers.

The vectors $x^*(y)$, $\hat{x}^*$, $z(x^*(y))$ and $v^*(x^*, \hat{x}^*)$ represent solutions to optimization problems which are directly or indirectly a function of the vector y, and therefore must be reoptimized any time y changes. The variables x^* and z^* are written as a function of y since they can be recalculated immediately following a change in y (technically speaking, the breakbulk flows z are directly a function of x and only indirectly a function of y). The variable v^* is written as a function of the LTL and TL flows x^* and $\hat{x}^*$ since the empties can be reoptimized only after these flows are determined. In our formulation, however, the truckload flows $\hat{x}^*$ are not a function of y or x^*; the optimal truckload flows are determined once, and do not need to be recalculated as long as the physical network of terminals and distances is unchanged.

The subproblems can be formulated as follows.

Routing Subproblem (RSP)

Decision variables $\{x, z\}$
Objective function

$$\min_{\{x,z\}} G_1(y, x) + G_2(z)$$

Subject to constraints 2, 3 and 4.

Decision	Level of user interaction	User control
Add/drop terminals	High	User initiated changes
Add/drop directs		Computer generated suggestions; user review and acceptance
Routing the LTL freight		Computer generated changes; user may review and override decisions
Routing truckload freight		No direct user control
Routing the empties	Low	No direct user control

Figure 4. Hierarchical structure of the decision variables.

The routing subproblem must be solved for a given vector of direct services y.

Truckload Freight Routing Problem (TRP)

Decision variables $\{\hat{x}_{ij}\}$
Objective function

$$\min_{\{\hat{x}\}} G_4(\hat{x}^*)$$

Subject to constraint 5.

Empty Balancing Subproblem (EBSP)

Decision variables $\{v\}$
Objective function

$$\min_{\{v\}} G_5(v^*(x^*, \hat{x}^*))$$

Subject to constraint 6.

Using this structure, any change in the vector y must consider the impact on all costs, as seen from **NDP**. The routing of the LTL freight, however, has been decoupled from the empty balancing costs, as evidenced by **RSP** and **EBSP**. Finally, while the flows of empties depend, in part, on the total flows of truckload freight, the routing of TL freight has been modeled independently of all other flows and, in fact, is independent of the vector y. Thus, **TRP** is the only subproblem that need not be reoptimized when y is changed.

At this point, the general structure of the solution approach should be clear. Using some well defined method for sequencing the search, the computer starts from a given load plan and attempts to add or drop individual links to and from the network. The method, described in Section 5, combines classical optimization concepts and user interaction. For each change to the network, the different subproblems must be solved to determine the impact of the change on the system. The next section briefly describes how these problems were solved.

4. Solving the Subproblems

The methodology for solving subproblems **RSP**, **EBSP** and **TRP** is outlined below, in the reverse order of their natural hierarchy.

The Empty Balancing Subproblem

As represented by **EBSP**, the routing of the empties is formulated as a classical linear transshipment problem with surpluses and deficits given by the vector $\{d\}$. Using an efficient network simplex code, the empty balancing problem could be solved to optimality for a network with 7,000 links in under 0.1 CPU seconds on an IBM 3081. During calibration tests, however, this approach was underestimating total empty movements by as much as 50%. The error arises in the failure to account for time-of-day and day-of-week effects, where empty miles are often generated to return drivers to their domicile (thus avoiding overnight expenses) or to get trailers in position at a particular point in time. For this reason, the calculation of empties was handled through a two phased approach. First, for every trailer going from A to B, a fraction w was assumed to return empty as a result of timing constraints. This approach created situations where empties might be going from A to B and from B to A, as actually occurs in practice. The movements of these *timing* empties were added to the surpluses and deficits created by the loaded movements (of TL and LTL freight), which were then input to a linear transshipment code. Calibration tests found that $w = 0.10$ produced total empty costs that agreed with the actual totals.

The Truckload Routing Problem

Truckload freight is included in the model only to reflect its effect on the trailer's surpluses and deficits at each terminal, which, in turn, determines the movement of empties. The routing of the freight is thus given by the shortest path over the line operations network. Total truckload costs will change, however, in strategic planning exercises that consider changes to the network itself.

The Routing Subproblem

By far the most important and difficult problem is the reoptimization of the routing of shipments following a change in the load planning network. The principal challenge here is speed of execution because it has to be done frequently and without compromising the value of the package as an interactive tool (if the process is too slow, the interactive feedback will be lost). For the size of the problem under consideration, it is unlikely that **RSP** could be reoptimized with sufficient speed to be useful. The difficulty is that it does not take advantage of the fact that at least 80% of the freight follows the least cost path (measured in terms of linehaul costs, c_{ij}, and handling costs, h_i). Such a least cost path solution is optimal if $F_{ij}(x_{ij}) = x_{ij}$ and is suboptimal only because $F_{ij}(x_{ij})$ is flat for $0 \leq x_{ij} \leq M_{ij}$.

The approach used in APOLLO is described in detail by Powell and Koskosidis (1986). It involves routing the freight along the least cost path subject to side constraints. These constraints, referred to as *routing overrides*, can be used to force flow over paths which are longer but that might have a lower marginal cost. The routine overrides can be viewed as decision variables that determine how freight is routed over the network. To describe this approach, define a routine override as

$$o_{ij}^s = \begin{cases} 1 & \text{if freight at } i \text{ destined for } s \\ & \quad \text{must next transfer at } j \\ 0 & \text{otherwise} \end{cases}.$$

If $\sum_j o_{ij}^s = 0$, then freight at i destined for s may follow the least cost path (in terms of linehaul plus handling costs). However, most freight is allowed to follow its *natural* (i.e., least linear cost) path, which is equivalent to saying that $\sum_j o_{ij}^s = 0$ for most terminals i and destinations s. The original routing subproblem can now be reformulated (see Appendix A3) as a two-tiered problem, where the first tier explicitly manipulates the set of overrides while the second tier, which is the workhorse of the entire model, involves solving the modified, shortest path problems very quickly.

The special structures of the load planning network and the routing subproblem make it possible to reoptimize the routing subproblem for a 300 terminal network in 0.05 and 0.2 CPU seconds on an IBM 3081. This fact means that the algorithm is fast enough to be used interactively.

5. Performing the Optimization

The natural approach for solving **NDP** is a local improvement heuristic that seeks to add or drop links to or from the load planning network, retaining changes that produce an overall reduction in total system costs. A batch model approach suffers from the major limitation that the computer may violate difficult to quantify constraints and, most importantly, must *believe its own numbers*. The alternative used by APOLLO allows the user to pose a series of what-if questions regarding adding or dropping links. The computer then calculates the impact of each possible change by reoptimizing the subproblems. The problem is that while the ability to pose what-if questions is important and provided for in APOLLO, an optimization process that depends completely on user initiated suggestions would fail due to the sheer size of the load planning network. It is impossible to expect the user to search exhaustively among the 15,000 possible places where links might be added or dropped.

APOLLO avoids this problem by suggesting network design improvements. These suggestions are generated by batch heuristics for network design similar to those suggested by Powell and Sheffi (1983) and extended in Powell. Instead of performing the iterative search, however, where the computer implements each change to reduce total costs (as in a batch approach), APOLLO simply stores all changes that produce a reduction in total system costs. These changes are later presented to the analyst in a list of *suggestions* for network improvement. The search itself is guided by the user who can specify, in a convenient fashion, regions of the network to look for improvements. Thus, the user may look for link additions from terminals in the northeast to the rest of the country, making use of symbols which define sets of terminals. With this, the user can focus on a particular part of the network or on a certain type of service.

The number of possible additions or deletions of links, as defined by a particular pair of terminal sets, could be extremely large. It was, therefore, necessary to develop an intelligent sequencing logic to screen out changes that were unlikely to produce a reduction in total costs. When dropping links, Powell and Sheffi (1983) found that sorting the links (connecting the origin's terminal set to the destination set) on the basis of least flow produced an efficient screen for identifying directs to be dropped.

The screening logic for identifying additions is somewhat more complicated, and was especially important given the large number of potential links that can be added. The method essentially developed a set of approximate costs, denoted w_{ij}, which represented the approximate change in costs from adding a direct from i to j. These values could be estimated for up to 15,000 links in 10 or 20 seconds. Once these savings approximations were calculated, the list of all possible link additions was sorted on the basis of these savings.

Given the ranking (for adding or dropping links), each of the top ranked links is individually added or dropped, and the change in total costs is calculated and stored. The user controls how far down the list the model proceeds before terminating the search. Each individual *add* or *drop* requires reoptimizing **RSP** and **EBSP**; after evaluating the change, the network is restored to its original form before testing the next link. Extensive numerical experiments show that the sequencing logic used for both adding and dropping links is quite efficient, with a very high success rate in the initial stages of the search. Experiments reported in Powell show over a 90% success rate in identifying link additions which show actual cost savings after the first 120 attempts, as opposed to a 5% success rate when links were tested randomly. This high efficiency in identifying good options is

No.		FROM	TO	FLOW	CHANGES IN COSTS LINE	VAR	H-BB	H-EOL	MPTY	TOTAL
# 1	ADD	ABR	CTW	2.14	-984	244	-79	0	-105	-924
# 2	ADD	RSW	WSH	1.96	-722	-55	-22	90	-130	-839
# 3	ADD	JAX	RSW	3.04	-1134	340	-120	86	334	-494
# 4	ADD	ABR	HAG	2.15	-460	105	-60	0	-45	-460
# 5	ADD	DMI	FTW	1.77	-475	230	-30	0	65	-210

ENTER: C (CONTINUE) T (TOP) I (IMPLEMENT) F (FORCE) D (DET. ANAL)
P (PRIMARY BREAK) E (ERASE) R (REFRESH TOP N)

Figure 5. Illustration of screen for reviewing and implementing suggested changes.

important because the analyst is sitting at the terminal while this search is being conducted.

After the search is completed, the model compiles a list of all the attempted changes, sorted in order of actual cost savings. This list is then presented to the user as a set of suggestions similar to that shown in Figure 5. For each addition, the screen shows how much flow would be attracted to the link if it were added, and the impact on each of the five cost categories. These categories include:

LINE linehaul cost, defined as

$$\sum_{i,j} c_{ij} x_{ij},$$

VAR linehaul variance, defined as

$$\sum_{i,j} (F_{ij}(x_{ij}) - x_{ij}) c_{ij},$$

which can be thought of as the cost of moving air,

H-BB handling costs at breakbulks,

H-EOL handling costs at origin end-of-lines,

MPTY cost of optimally rebalancing the empty trailers.

The distinction between linehaul cost and variance may seem unusual, since the sum of these two numbers gives the total transportation costs. This breakdown, however, is a way of distinguishing between hard and soft numbers. The total linehaul cost, designated as LINE, is a very hard and reliable number, whereas the model's estimate of the additional cost of running partially loaded trailers can be somewhat soft. The reason is that in the field trailers are usually filled in creative ways, in part, by using truckload freight. Thus, a recommendation that shows an increase in linehaul cost, and a large reduction in linehaul variance, might be suspect.

At the bottom of the screen in Figure 5 is a short menu that allows the user to implement and edit specific changes. For example, suggestions 1, 4 and 5 are implemented by entering "I 1 4 5" on one line. Of particular importance is option *D*, which allows the user to perform a detailed analysis of any specific suggestion. This option puts the user into a separate submenu for a complete and detailed summary of what changes as a result of adding or dropping the link in question.

6. Architecture of the Model

The logic just described was implemented in the APOLLO package, which is currently used on an ongoing basis by PIE. With over 40,000 lines of code and a 300 page user's manual, it is impossible to describe the software in any detail. A brief summary of the program, its menus, the reports and the graphics capabilities are given in Powell and Sheffi (1986). It is useful, however, to highlight important aspects of the architecture of the program that proved to be particularly helpful in the implementation. Figure 6 gives an overview of the program, summarizing the major menus, some of the options, and the movements between the menus. The major components of the program include the following.

Data input. The software reads the current load plan, the previous month's freight bill file and the network distances from existing data sets. Overrides are automatically inserted when the model determines that freight is not currently flowing over the least cost path. LTL and TL origin-destination matrices are developed from the freight bill file. No preprocessing is required of these data sets.

Data editing routines. Data can be interactively modified by the user, including dozens of terminal parameters, adding/dropping/changing the status of a terminal, manipulating the freight flows in a variety of flexible ways (for strategic planning purposes), and modifying the line operations network.

The optimization routines. There are over 10,000 lines of code for finding the shortest path subject to

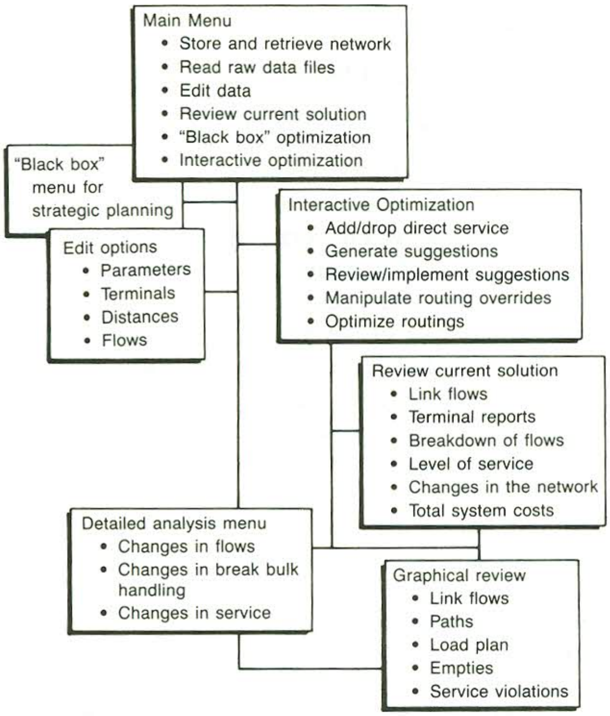

Figure 6. Overview of the structure of the software package.

overrides, reoptimizing the routing subproblem (**RSP2**), optimizing the overrides themselves (**RSP1**), balancing empty trailers (**EBSP**), and guiding the search sequence.

Interactive optimization routines. These are the routines that handle the interface between the user and the optimization routines. Included is the capability to let the user guide the computer as to what types of changes to look for (e.g., adding/dropping links, manipulating overrides) and where to look (between which sets of terminals). After compiling a list of changes, these routines present the list of suggestions, allowing the user to accept or reject changes as well as to conduct a detailed analysis of any given change.

The "black box". A set of routines for optimizing a network with limited interactive control. These would be used in strategic planning exercises.

The graphics routines. These show flows on directs between user specified groups of terminals, and only those directs with flow in a given range. Other reports show the load plan into a destination, the shortest path over the line operations network into a destination, the shortest path over the line operations network into a destination, service violations, empty flows, truckload flows, and the changes in flows due to adding or dropping a given direct service. Complete windowing capabilities are provided.

Interactive (screen) and printed reports. These include system statistics (with total costs broken down into ten categories, load average and fleet requirements), flows into and out of a terminal, level of service between groups of terminals, detailed path descriptions between terminals, a summary of the flow on a direct service, a detailed breakdown of which traffic lanes use a particular direct service, and a detailed analysis of the level of service between two terminals. Particularly important are a set of routines for performing detailed comparisons between any two load plans. When compared against a base case load plan, these reports show what has changed during a complete load planning session.

File handling routines. The user may save up to 20 different scenarios, interactively review what networks have been saved and load any network into the core. This helps to minimize contact with the MVS/TSO operating system, which is standard on most corporate IBM mainframes (and which often baffles all but the most determined analysts).

Other special characteristics of the software include: a shallow menu structure to eliminate the user from becoming lost in the model; command stacking to allow the simultaneous stacking of numerous commands that speed performance for the experienced user; complete robustness to user input errors; and consistent use of vocabulary familiar to PIE.

7. Issues in Implementation

Up to now, the discussion has focused on the effects of real-world constraints on model formulation, solution approaches and the software architecture. Paramount to this discussion is the use of an interactive optimization approach to handle unquantifiable constraints and tradeoffs. The project involved, however, other issues that played an important role in the successful implementation of APOLLO. These include the following.

The environment. The problems of managing a growing carrier, increasing frustration with the prior planning methodology, and the pressure in the newly deregulated market to control costs while maintaining service, all acted as strong incentives to try something new.

Compatibility with the existing planning process. APOLLO runs off existing data sets. After an optimization is complete, APOLLO produces a data set that details changes to the load plan in the same format produced under the old process. In addition, the ability to document "what changed" enables the same management review procedures used before to be used with APOLLO. In short, the changes were largely confined to the Industrial Engineering Department.

Organizational independence from the MIS department. The MIS functions at PIE are handled by a subsidiary, Trans-Tel Inc., located several miles away. With APOLLO, the IE staff was furnished with a computer terminal and direct access to the main computer for the first time, providing a level of freedom never enjoyed before. Significantly, the MIS group was very supportive of APOLLO.

Capabilities of the MIS group. Trans-Tel provided a good computer with quality support, and could supply all the fundamental data required by the model. The fact that the load plan was already computerized (unusual in the trucking industry) significantly simplified the implementation of APOLLO's recommendations.

Developing a new load planning process. The method used prior to APOLLO, though unwieldy, was straightforward and well understood. During the implementation of APOLLO, it quickly became clear that an entirely new load planning process, starting with loading data and ending with communicating the plan to the field, was needed. This process included a new set of mechanics for identifying directs to be added and dropped, modifying overrides and reviewing the final solution for reasonableness and feasibility. The flexibility of the package allowed the IE staff at PIE to develop such a new process largely on their own.

The structure of the problem. Central to the successful use of the computer was the presence of the routing subproblem, which was extremely large but highly amenable, to a computer solution using classical techniques. Simple heuristics were not adequate.

Keeping the principles simple. Despite the large size of the package, the basic principles driving every component of the system are extremely simple. The costing logic, routing of the freight, handling of level of service, and determining the system impacts of a change are all conceptually simple. This simplicity derives, in large part, from letting the user account for certain approximations (work rules, day of week effects, level of service) and by simplifying complex tradeoffs, such as the decoupling of the routing subproblem (**RSP**) and the empty balancing subproblem (**EBSP**). As a result, it is possible to document how the model calculates its savings.

8. The Results

The impact of APOLLO on PIE can be viewed in two ways: immediate cost reductions due to an improved load plan, and a fundamental change in the PIE approach to network planning. Given the broad impact on basic planning practices, coming at a time when the carrier was undergoing tremendous changes, it is difficult to quantify the cost savings with any accuracy. Two exercises, however, shed light on the potential impact of the model. Immediately prior to the first full implementation of the model, APOLLO was used in a study to locate a new breakbulk in the northeast. The Industrial Engineering staff had already performed a manual analysis which focused on distances from the break to its satellites in the northeast without any corresponding network effects. Their study required approximately one month to analyze a single location and produced estimated savings of \$0.5 million annually. Using APOLLO, eight separate locations were analyzed in depth, requiring approximately one week of continuous work. This study showed that the first location studied manually by PIE would actually produce losses of \$1.0 million, and recommended a different location, which would save \$2.0 million annually in operating costs.

Following this study, the model was implemented at the carrier's headquarters, and an intensive load planning effort was undertaken to improve the current load plan. Starting with the existing load plan, an intensive two week exercise was undertaken to optimize the load plan, and to identify an estimated \$7–10 million annually in transportation and handling costs (out of total operating costs of \$400 million). Within the planning team, these were always referred to as "APOLLO" savings, since it was not clear to what extent the plan inside the computer was actually being followed in the field. It was the feeling of the authors that continued use of the model to identify savings, and the secondary benefits derived from a cleaner load plan, at least matched these initial savings, but no formal study was attempted by PIE to rigorously estimate these savings. Even with such numbers, the \$500,000 development cost of APOLLO was certainly justified, but there were strong reasons to suspect that the true value of the model was being underestimated.

The best evaluation of APOLLO is in the effect it has had on fundamental planning and operating practices. Mr. Robert Radford, head of the Industrial Engineering Department at PIE, summarized five major areas where changes had occurred:

Changing operating philosophy. APOLLO helped change a basic philosophy of splitting freight at the origin end-of-line in order to fill trailers headed as close to the destination as possible. By modeling the higher handling costs at origin from splitting freight, PIE significantly reduced the number of directs out of end-of-lines going to breaks other than the primary break.

Routing according to marginal cost. Due in part to union work rules, the marginal cost of handling additional freight at a breakbulk can be significantly different than the average cost. APOLLO is now given the marginal handling costs for each breakbulk so that the routing of freight through the breakbulks is performed on the basis of marginal rather than average costs.

Greater willingness to test new ideas in the field. APOLLO's ability to quickly redesign a load plan allows the carrier to test new concepts in the field, and to respond rapidly depending on the outcome. A recent experiment was run where a new operating policy was implemented at each terminal, which governed when the terminal managers could dispatch trailers. The concept was to reduce specific driver expenses generally incurred between the primary break and its satellites, which are not accurately modeled inside APOLLO. When the experiment failed, a new load plan was designed and implemented in two weeks. Such trial and error testing is needed to develop new ideas, and APOLLO gives management the opportunity to test major changes and then respond quickly depending on the outcome.

Expanding use of different options. APOLLO allows rapid analysis of new transportation options now available in a deregulated environment, including doubles, triples, rail and cargo ship. The network model quickly identifies which shipments could use a new option and summarizes the total system impact of the change. As a result, PIE explores and uses these possibilities much more aggressively than it would otherwise.

Improved load plan enforcement. Prior to APOLLO, and due in part to the recent merger of PIE and Ryder Truck Lines, the load plan in place was difficult for terminal managers to follow due to certain inconsistencies between the plan and the manager's incentive system. As a result, terminal managers in the field were making many important routing decisions without understanding the system impacts of their actions, a practice which encouraged greater independence on the part of field managers. This decision making showed up in PIE's "misload" report, which depicts shipments that do not follow the load plan. In the last year of using APOLLO, the percentage of misloads dropped from 8 to 4%; the misload statistic when APOLLO was first implemented (two years ago) was not kept, but was even higher. The value of APOLLO is that it restored the confidence of field managers in the plan being developed in the central office, thereby improving overall enforcement in the field.

At the end of the project, in the latter half of 1985, it appeared that this paper would have to conclude that "the operation was a success but the patient died." After beginning the project with Ryder Truck Lines, IU merged Ryder with PIE to form Ryder/PIE (after which the name was changed to PIE). The planning of a new network was handled manually, and the resulting company was only marginally profitable (for a variety of reasons). One year later, the combined carrier was merged with a bankrupt regional carrier sending PIE into the red with annual losses of over $80 million at one point. The substantial savings identified by APOLLO were dwarfed by comparison, and the many changes implemented to reduce these losses made it impossible to document the direct effects of APOLLO.

At the time of this writing (September 1988) it appears that PIE will emerge as a profitable national LTL motor carrier. Following a difficult contraction period, PIE is becoming increasingly profitable and is again beginning to grow. If this continues, its success will be attributed to the steady process of refining basic day-to-day operations in which APOLLO continues to play an important role. In the meantime, the system has been adopted by what is now the nation's largest LTL carrier, where it is being aggressively integrated into the tactical and strategic planning process. We continue to find that as new carriers adopt the system, they learn an entirely new perspective for understanding network operations, simply by virtue of using an optimization based system that quickly and interactively gives total cost impacts of different suggestions.

Appendix

A1. The Network Design Problem

The load planning problem can be posed as a large mathematical programming problem. The following notation is used:

B = set of all breakbulk terminals,
E = set of all end-of-line terminals,
LO = set of links in the line operations network,

LP = set of all potential links in the load planning network,

LN = set of links in the load planning network over which direct service is being offered, where LN < LP.

For the purposes of this description, assume that LO ⊆ LP. Since drivers generally cannot relay at end-of-lines, the line operations network, like the load planning network, includes only links to and from breakbulks. Also, the line operations network includes additional nodes representing driver relay points.

The problem parameters include:

c_{ij} = the linehaul cost per trailer (loaded or empty) from i to j, $(i, j) \in$ LO,

M_{ij} = the minimum frequency (trailers per week) from i to j, $(i, j) \in$ LP,

q_{rs} = LTL flow (trailerloads per week) originating at terminal r and destined for terminal s, $r, s, \in \mathrm{E} \cup \mathrm{B}$,

$q_{r\cdot}$ = the total flow from origin r to all destinations,

$$= \sum_s q_{rs},$$

$q_{\cdot s}$ = the total flow from all origins to destination s,

$$= \sum_r q_{rs},$$

$\hat{q}_{rs}$ = the total truckload (TL) freight from origin r to destination s,

h_i = the handling cost per trailer at i, $i \in$ B,

$g_i(N_i)$ = the handling cost per trailer at origin i, $i \in$ E, given N_i outbound load planning links from i (N_i is defined below),

S_i = the set of satellites of breakbulk i, $i \in$ B.

For a link, $(i, j) \in$ LP but $(i, j) \notin$ LO, c_{ij} is defined as the shortest path cost from i to j over the line operations network.

The decision and other related variables in the problem formulation include

$$y_{ij} = \begin{cases} 1 & (i, j) \in \mathrm{LN} \\ 0 & \text{otherwise} \end{cases}$$

$$p^s_{ij} = \begin{cases} 1 & \text{if flow at } i \text{ destined for } s \text{ must move next to } j, \quad (i, j) \in \mathrm{LN} \\ 0 & \text{otherwise} \end{cases}$$

R = the set of all permissible routing vectors $\{p^s_{ij}\}$ that satisfy the cluster constraint,

x^s_{ij} = LTL flow on link (i, j) with destination s,

x_{ij} = the total LTL flow on link (i, j),

$\hat{x}^s_{ij}$ = TL flow on link (i, j) with destination s,

$\hat{x}_{ij}$ = the total TL flow on link (i, j),

d_i = the net supply of or demand for trailers at i produced by the flows of loaded LTL and TL trailers,

N_i = the number of outbound load planning links from i,

v_{ij} = the flow of empty trailers from i to j, $(i, j) \in$ LO.

For simplicity, all flows are assumed to be in units of 45 foot trailers, with one driver and one tractor per trailer movement. The use of twin 28 foot trailers behind one tractor, which has become prevalent in the last few years, complicates the presentation without contributing to the understanding of the problem. (Of course, the load planning software developed for PIE had to be designed to handle "doubles.") It is also assumed that truckload freight moves on full trailers, even if the flow of freight is fractional. This assumption is justified because the flow averaging used to calculate $\{\hat{q}_{rs}\}$ is not representative of the infrequent TL shipments.

The LTL network design problem can be stated as

$$\min_{\{y,p,v\}} \sum_{ij \in \mathrm{LP}} c_{ij} F_{ij}(x_{ij}) y_{ij} + \sum_{i \in \mathrm{B}} h_i z_i + \sum_{i \in \mathrm{E}} g_i(N_i) q_i + \sum_{ij \in \mathrm{LO}} c_{ij} v_{ij} + \sum_{ij \in \mathrm{LO}} \hat{x}_{ij} c_{ij} \tag{A1}$$

subject to

integrality of the network design variables:

$$y_{ij} = (0, 1) \quad \text{for any } i, j \tag{A2}$$

routing constraints:

$$p^s_{ij} = (0, 1) \quad \text{for any } i, j, s \tag{A3}$$

$$\sum_j p^s_{ij} = 1 \quad \text{for any } i, s \tag{A4}$$

$$p^s_{ij} \leq y_{ij} \quad \text{for any } i, j, s \tag{A5}$$

$$\{p^s_{ij}\} \in R \tag{A6}$$

definition of breakbulk flows:

$$z_i = \sum_k x_{ki} - q_{\cdot i} \quad \text{for any } i \tag{A7}$$

LTL link flows:

$$x^s_{ij} = \left[q_{is} + \sum_k x^s_{ki} \right] p^s_{ij} \quad \text{for any } i, j, s \tag{A8}$$

$$x_{ij} = \sum_s x^s_{ij} \quad \text{for any } i, j \tag{A9}$$

Net surplus/deficit of trailers due to LTL and TL loaded moves:

$$d_i = \sum_k F_{ki}(x_{ki}) + \sum_r \hat{q}_{ri} - \sum_j F_{ij}(x_{ij}) - \sum_s \hat{q}_{is} \quad \text{for any } i \quad \text{(A10)}$$

Empty balancing:

$$\sum_j v_{ij} - \sum_k v_{ki} = d_i \quad \text{for any } i \quad \text{(A11)}$$

TL link flows:

$$\sum_j \hat{x}^s_{ij} - \sum_k \hat{x}^s_{ki} = \begin{cases} \hat{q}_{is} & i \neq s \\ -\sum_r \hat{q}_{ri} & i = s \end{cases} \quad \text{(A12)}$$

$$\hat{x}_{ij} = \sum_s \hat{x}^s_{ij} \quad \text{for any } i, j. \quad \text{(A13)}$$

A2. The Decomposition Approach

The optimization problem **NDP** can be broken down into a series of subproblems using the natural hierarchy of the problem. The highest level, the network design problem, takes into account all the cost categories. This level treats only the vector y as decision variables and treats all the other variables implicitly as functions of y.

Network Design Problem (NDP)

Decision variables $\{y_{ij}\}$

Objective function

$$\min_{\{y\}} \sum_{ij \in \mathrm{LP}} c_{ij} F_{ij}(x^*_{ij}(y)) y_{ij} + \sum_{i \in \mathrm{B}} h_i z^*_i(y) + \sum_{i \in \mathrm{E}} g_i(N_i(y)) q_i + \sum_{ij \in \mathrm{LO}} c_{ij} v^*_{ij}(x^*) + \sum_{ij \in \mathrm{LO}} c_{ij} \hat{x}^*_{ij} \quad \text{(A14)}$$

Constraints (A2).

In problem **NDP**

$y = (\ldots, y_{ij}, \ldots)$ is the vector of decision variables y_{ij},

$x^*_{ij}(y)$ = optimal set of LTL link flows for a given load planning network, defined by y,

$x^* = (\ldots, {}^*_{ij}(y), \ldots)$

$z^*_i(y)$ = optimal flow handled at breakbulk i for a given set of link flows $x^* = x^*(y)$,

$v^*_{ij}(x^*)$ = optimal flow of empties on link (i, j) given x^*,

$\hat{x}^*_{ij}$ = optimal flow of TL freight on link (i, j).

The vectors $x^*(y)$, $z^*(y)$, $v^*(x^*)$ and $\hat{x}^*$ represent solutions to optimization problems that are directly or indirectly a function of y, and therefore, must be reoptimized any time y changes. The variables x^* and z^* are written as a function of y because they can be recalculated immediately following a change in y. The variable v^* is written as a function of x^* and $\hat{x}^*$ since the empties can be reoptimized only after the TL and LTL flows have been rerouted. Under the modeling assumptions made here, $\hat{x}^*$ is not a function of either y or x^*.

The subproblems can be formulated as follows.

Routing Subproblem (RSP)

Decision variables $\{p^s_{ij}\}$

Objective function

$$\min_{\{p^s_{ij}\}} \sum_{ij \in \mathrm{LN}} c_{ij} F_{ij}(x_{ij}) + \sum_{i \in \mathrm{B}} h_i z_i \quad \text{(A15)}$$

Constraints (A3)–(A9).

Truckload Freight Routing Problem (TRP)

Decision variables $\{\hat{x}_{ij}\}$

Objective function

$$\min_{\{\hat{x}_{ij}\}} \sum_{ij \in \mathrm{LO}} c_{ij} \hat{x}_{ij} \quad \text{(A16)}$$

Constraints (A12) and (A13).

Empty Balancing Subproblem (EBSP)

Decision variables $\{v_{ij}\}$

Objective function

$$\min_{\{v_{ij}\}} \sum_{ij \in \mathrm{LO}} c_{ij} v_{ij} \quad \text{(A17)}$$

Constraints (A10) and (A11).

A3. The Routing Subproblem

RSP1

$$\min_{\{o^s_{ij}\}} \sum_{ij \in \mathrm{LN}} c_{ij} F_{ij}(x^*_{ij}(o)) + \sum_{i \in \mathrm{B}} h_i z^*_i(o) \quad \text{(A18)}$$

subject to

$$o^s_{ij} \leq y_{ij} \quad \text{for any } i, j \quad \text{(A19)}$$

$$\sum_j o^s_{ij} \leq 1 \quad \text{for any } i \quad \text{(A20)}$$

$$o^s_{ij} = (0, 1) \quad \text{for any } i, j.$$

The vectors $x^*(o)$ and $z^*(o)$ are the optimal solutions to **RSP2**.

RSP2

$$\min_{\{p_{ij}^s\}} \sum_{ij \in \text{LN}} c_{ij} x_{ij} + \sum_{i \in \text{B}} h_i z_i \tag{A21}$$

subject to (4)–(10) and

$$p_{ij}^s \geqslant o_{ij}^s. \tag{A22}$$

Note that for both **RSP1** and **RSP2**, the summation of linehaul costs is performed over the set **LN** rather than **LP**. This is done because the routing subproblem is defined necessarily over the set of links currently in the load planning network.

Without (A22), **RSP2** could be solved as a series of shortest path problems into each destination. As stated, **RSP2** can be solved optimally using the modified label correcting algorithm in Powell and Koskosidis (1986). A source of significant complexity in this algorithm is the presence of cluster constraints. Finding the shortest path subject to these constraints is a moderately difficult optimization problem, but the algorithm developed appears to be very fast.

It is possible to design an algorithm to optimize the overrides. In the context of the local improvement heuristic for **NDP**, however, the idea is generally not to optimize **RSP1** following a change in y, but rather to reoptimize only **RSP2**. Such an approach will generally give a very good approximation of $x^*(y)$, with one exception. If, after adding a link, the total flow attracted to the new link after reoptimizing **RSP2** is less than the minimum, **RSP1** has to be optimized further. This can be accomplished by manipulating overrides to attract additional flow over the new link. These algorithms are described in depth in Powell (1986). The impact of adding a link, then, might require optimizing first **RSP2** and then, where necessary, refining the solution by further optimizing **RSP1**. Note that every change in the set of overrides requires a corresponding reoptimization of **RSP2**. If the override being changed is on freight destined to an end-of-line, then only one minimum path tree calculation is required. If the override is on freight into a breakbulk, then a shortest path tree calculation is required for freight into the breakbulk as well as, potentially, for freight into each of the satellites.

Acknowledgment

A project such as this simply cannot take place without the help of a number of people. First is John Terry, at the time the head of the Land Transportation Group at IU International, who made the decision to fund the project, despite virtually no track record on the part of the authors in the trucking industry. Dr. Brian Kullman, then on John Terry's staff, played the important role of project monitor, particularly in the early stages when there was little involvement with Ryder Truck Lines. Al Pinkerton, who succeeded Dr. Kullman, continued the enthusiastic support of the project.

The true heroes of this project are the Industrial Engineering Department at PIE who undertook the enormous time commitment to educate the authors in the trucking industry and guide the development of APOLLO to ensure that it met the needs of the real world. Many subtle modeling issues arose during these discussions that often had a direct impact on our analytical and algorithmic approaches. Bob Stewart, head of the IE Department at the beginning of the project, Mark Johnson, Dennis Johnson and Ralph White warmly accepted our often misguided efforts at helping them do their job, and with tremendous patience tested early versions of the package. Special thanks go to Paul Bodenstein who was the first "APOLLO expert" at PIE, responsible for actually operating the code. Paul demonstrated the patience of Job in operating the initial version of the model, which was clumsy and not entirely bug free, in addition to being very complex. Bob Radford, the current head of the IE department, and Steve Hassler, who succeeded Paul as the "APOLLO expert" have shown equal support in continuing to use APOLLO, and, in particular, their creative application of the model to a range of problems which the model was not originally intended to handle.

References

BALAKRISHNAN, A. 1984. Valid Inequalities and Algorithms for the Network Design Problem with an Application to LTL Consolidation. Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge.

BALAKRISHNAN, A., AND S. C. GRAVES. 1985. A Composite Algorithm for the Concave-Cost LTL Consolidation. OR Center Working Paper, Sloan School of Management, MIT, Cambridge.

BARKER, H. H., E. M. SHARON AND D. K. SEN. 1981. From Freight Flows and Cost Patterns to Greater Profitability and Better Service for a Motor Carrier. *Interfaces* **11,** 4–20.

LAMAR, B. W., AND Y. SHEFFI. 1985. Optimal Network Design for LTL Trucking. Working Paper CTS 85-7, Center for Transportation Studies, MIT, Cambridge.

LAMAR, B. W., Y. SHEFFI AND W. B. POWELL. 1984. A Lower Bound for Uncapacitated Multicommodity

Fixed Charge Network Design Problems. OR Center Working Paper, Center for Transportation Studies, MIT, Cambridge.

MAGNANTI, T. L., AND R. T. WONG. 1984. A Dual Ascent Approach for Network Design Problems. OR Center Working Paper, Sloan School of Management, MIT, Cambridge.

Multisystems. 1979. Description of Motor Carrier Breakbulk Planning Model: Program User's Manual. MULTISYSTEMS, Cambridge, Mass.

POWELL, W. B. 1986. A Local Improvement Heuristic for the Design of Less-Than-Truckload Motor Carrier Networks. *Trans. Sci.* **20,** 246–257.

POWELL, W. B., AND I. KOSKOSIDIS. 1984. Shipment Routing Algorithms with Tree Constraints: Relaxation Methods. Working Paper EES-84-8, Department of Civil Engineering, Princeton University, Princeton, N.J.

POWELL, W. B., AND I. KOSKOSIDIS. 1986. Shipment Routing Algorithms with Tree Constraints: Shortest Path with Side Constraints. Working Paper EES-86-4, Department of Civil Engineering, Princeton University, Princeton, N.J.

POWELL, W. B., AND Y. SHEFFI. 1983. The Load Planning Problem of Motor Carriers: Problem Description and Proposed Solutuion Approach. *Trans. Sci.* **17,** 471–480.

POWELL, W. B., AND Y. SHEFFI. 1986. Interactive Optimization for Motor Carrier Load Planning. *J. Logist.* **7,** 64–90.

ROY, J. 1984. Un Modèle de Planification Globale pour le Transport Routier des Marchandises. Publication 402, Centre de Recherche sur les Transports, Université de Montréal, Montréal, Canada.

Temple, Barker and Sloane. 1978. Motor Carrier Linehaul and Breakbulk Planning Model, Version III, Program User's Manual. Temple, Barker and Sloane, Inc., Lexington, Mass.