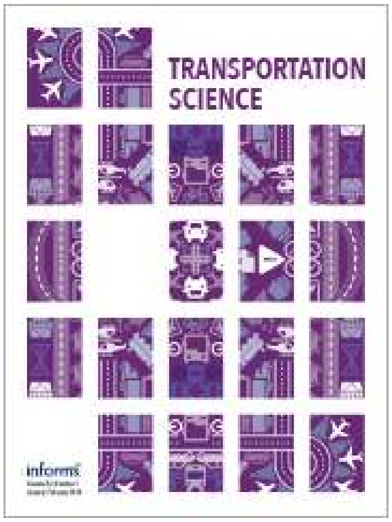

Transportation Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

The 1990 Transportation Science Section Dissertation Prize Competition

Warren B. Powell,

TSS Dissertation Abstracts

The 1990 Transportation Science Section Dissertation Prize Competition

WARREN B. POWELL

Chairman, 1990 TSS Dissertation Prize Competition

The 1990 Transportation Science Section Dissertation Prize Competition received seven nominations from six universities, with research topics ranging from a queueing network analysis of automated handling systems to primal partitioning solutions for the routing of shipments for less-than-truckload motor carriers. As in previous years, the judging committee was impressed by both the technical depth and richness of the presentations, as well as the breadth and creativity displayed in the choice of applications.

This year, the winner is Jeffrey I. McGill for his dissertation, "Optimization and Estimation Problems in Airline Yield Management." The committee was impressed by the organization and clarity of the work, the strong methodological contribution, and the importance of the problem area to the transportation community. The research addressed a rich optimization problem with the statistical estimation challenges involved in working with actual observations. Perhaps one of the most important contributions of the thesis is that it helps to provide a strong scientific foundation to a difficult problem area.

The judging committee was comprised of Marc Goetschalckx, Georgia Institute of Technology; Gilbert Laporte, Universite de Montreal; Hani Mahmassani, University of Texas at Austin; Anna Nagurney, University of Massachusetts at Amherst; Warren Powell, Princeton University (chair); and Richard Wong, AT&T Bell Laboratories.

In view of the high quality of all the dissertations that were nominated for this year's competition, we are including the extended abstract of each thesis. Separate from providing some much deserved recognition, we hope that these abstracts help to provide a snapshot of ongoing research programs around the country.

Optimization and Estimation Problems in Airline Yield Management

by

JEFFREY I. McGILL

One of the obvious impacts of the deregulation of North American airlines that occurred in the 1970s has been increased price competition and the resulting proliferation of discount fare booking classes. The problem of optimal allocation of available seating capacity to these new fare classes has become of paramount importance to air passenger carriers. Under the general headings "Yield Management," "Revenue Control" or "Airline Seat Inventory Control," this and related stochastic optimization problems have become central to airline tactical planning efforts. In recent years, decision-makers in other transportation modes and in sectors of the entertainment and hotel/motel industries have become interested in similar problems.

Taken as a whole, the yield management problem is extremely complex—so much so that no single model formulation has come close to encompassing all facets of the problem. The objective of this dissertation is to examine in isolation four important components of the problem with the aim of obtaining insights into the nature of good solutions. Such solutions can be useful as inputs to more comprehensive treatments and, taken by themselves, as good but not necessarily optimal partial solutions to the overall problem.

The introductory chapter of this dissertation discusses the components of the yield management problem,

Transportation Science
Vol. 25, No. 3, August 1991

0041-1655/91/2503-0245 $01.25

surveys previous research, and highlights the particular sub-problems that are the subject of the dissertation.

Chapter 2 of the dissertation deals with the problem of assigning seats to multiple, stochastically independent fare classes that are booked in a nested, sequential fashion into a common seating pool on one leg of a flight. It is shown that a fixed-limit booking policy that maximizes expected revenue can be characterized by a simple set of conditions on the subdifferential of the expected revenue function. These conditions are transformed to a set of conditions relating the probability distributions of demand for the various fare classes to their respective fares. Characterization of the problem as a series of monotone optimal stopping problems proves optimality of the fixed-limit policy over all admissible policies. A comparison is made of the optimal solutions with the approximate solutions obtained by P. Belobaba using the expected marginal seat revenue (EMSR) method.

Chapter 3 introduces a seat allocation model that allows for stochastic dependence between two fare classes. A well known two-fare seat allotment formula, which requires the independence assumption, is generalized to a condition which requires only a much weaker and more realistic monotonic association assumption. An extension of the model allows for control of full fare passenger spillage (rejection of full fare reservation requests when a flight is fully booked) and consideration of the impact of passenger goodwill on seat allocation decisions. In addition, the model is used to provide a rigorous proof of a formula (proposed by others) for optimal seat allocations in the special case that dependency arises because of upgrades.

Chapter 4 examines a basic version of the problem that has received the most attention from airline researchers over the years—that of determining suitable overbooking levels in one fare class. This analysis is motivated by the recognition that the seat allocation problem for two dependent fare classes and the overbooking problem for a single fare class have very similar structures. This chapter exploits this similarity in developing a simple model for the overbooking problem. Specification of the passenger confirmation process as a Bernoulli process leads to a simple condition for an optimal overbooking level that is very similar in structure to the simple two-fare allotment rule mentioned above. This condition is further simplified with approximation methods, and it is shown that under certain circumstances the simple ratio of cabin capacity to the confirmation probability gives a good estimate of the optimal overbooking level.

The two-class, dependent demand booking rule obtained in Chapter 3 can be easily implemented if the joint probability distribution of demands for the two fare classes can be estimated. This would be a routine problem of estimating a two-equation multiple regression model on the basis of historical demand data were it not for the fact that demand occurring after a capacity limit is reached cannot be recorded. That is, the historical data are *censored* by the presence of booking limits and the finite capacity of aircraft. It is well-known that failure to take into account such censorship can lead to badly biased estimators and unreliable forecasts. Chapter 5 of the dissertation deals with the problem of estimating a demand model on the basis of censored data for two or more correlated classes of demand that are subject to a common capacity constraint. It is shown that an adaptation of the *EM method* of Dempster, Laird and Rubin can provide maximum likelihood estimates of the parameters of the demand model under these circumstances. The results of numerical trials of the method demonstrate that the technique is both accurate and computationally efficient on problems of realistic size.

This dissertation was completed at the University of British Columbia under the supervision of Prof. Shelby Brumelle.

A Primal Partitioning Solution For Multicommodity Network Flow Problems

by

JUDITH M. FARVOLDEN

The motivating application of this research is the freight distribution and scheduling problem of Less-Than-Truckload (LTL) motor carriers. LTL operations are characterized by the consolidation of a number of small shipments to form a load which is then transported between two terminals in the operations network of the carrier en route to the destination terminal. The trend of shippers is toward pickup and delivery schedules which are coordinated with their production schedules. As they strive to reduce costs by lowering inventory levels, shippers are requiring carriers to deliver goods *just-in-time* for further processing. To meet these requirements, LTL carriers must consider the level of service accorded to each shipment, necessitating the incorporation of detailed scheduling into their load planning activities.

The movement of shipments between terminals can be characterized by the flow of commodities in a network. To model the detailed level of service requirements, a commodity in this model is defined as a single shipment with a single origin and a single destination, moving at a particular point in time, which allows us to specify a delivery standard for each shipment in the system. A mid-sized LTL carrier transports perhaps 15 thousand shipments daily. To capture the dynamic nature of the scheduling problem we formulate the problem as a multiperiod multicommodity network flow problem (MNFP) to optimally schedule shipments while explicitly accounting for level of service and trailer capacities.

The result is a very large-scale programming problem. The focus of the research is to find ways of reducing the solution time and of extending the capabilities of the primal simplex method to these ultra-large problems. While considerable effort has been devoted to the development of specialized solution techniques to solve MNFP, a review of the literature indicates that none of the techniques developed to date solves problems as large as those under consideration here.

We have developed, implemented and tested a new solution approach for solving MNFP which is based upon both primal partitioning and decomposition techniques. The partitioning of the path variables, which are the decision variables in this solution, was motivated by some arguments drawn on the mechanisms of the primal simplex method. First, it was noted that due to the extreme point nature of basic, feasible solutions, if there is adequate capacity on a path for a shipment, the entire shipment will be assigned. The shipment will be split only if an arc in the path becomes saturated and the associated slack variable nonbasic. Second, in each pivot path flows must be adjusted such that capacity constraints on saturated arcs remain binding and the flow conservation constraints are satisfied. Interactions occur among the binding capacity constraints and the flow conservation constraints for split shipments.

We note that in a system with adequate capacity only a small subset of the arc capacity constraints will be binding, conjecture that only a subset of the shipments will be split over multiple paths, and show that only variables associated with split shipments will change value in an iteration of the revised simplex method when the constraint on a saturated link is involved. The basis can be partitioned to isolate a working basis of greatly reduced dimension which contains all of these variables. All nonunimodular aspects of the full basis are contained in this working basis which itself displays special structure. The primal simplex method is specialized for application to this partitioned basis and the result is an elegant set of equations for the computations required.

A *pivot type* is characterized by the entering and exiting variables. All feasible combinations of entering and exiting variables are enumerated and the pivoting operations specialized for each case. By specializing the simplex method to consider each of the possible pivot types, the time required to search for the exiting variable can be reduced. The columns of the arc-chain incidence matrix are generated by a dual network simplex method for updating shortest paths when link costs change. Further time savings are possible through other shortcuts to the simplex method which exploit the underlying network structure. The majority of the simplex operations performed on the partitioned basis are simply additive and network operations. Only the operations on the sparse, near-triangular working basis are performed using a commercially available matrix factorization package. The result is a large (over 10,000 lines of Fortran) package of pure optimization code.

The partitioning used in this solution approach is significantly different than those used previously. The MNFP is most frequently formulated on link variables. The flow conservation constraints, partitioned to expose their block angular and network structure are viewed as the "easy" constraints and the link capacity constraints as the "complicating" constraints that make the problem difficult to solve. The capacity constraints may be further partitioned to isolate a working basis of rank equal to the number of saturated arcs. This type of partition ignores the interactions that occur among the binding capacity constraints and the flow conservation constraints for split shipments.

However, the PPLP solution has elements of other large scale optimization techniques. Specifically, it exploits the generalized upper bounding structure and is performed on a basis that is similar within a change of variables to the Dantzig-Wolfe Master Problem. In light of this equivalence, the PPLP solution can be viewed as a primal partitioning specialization of Dantzig-Wolfe decomposition for a class of MNFP. Alternatively, the algorithm is simply a specialized application of the revised simplex method with primal partitioning and implicit path enumeration.

We present the results of computational experiments undertaken with two types of data sets and three exact solution techniques: the PPLP code, Dantzig-Wolfe decomposition and a straightforward solution by a commercial LP solver (MINOS). The problems of the first data set are randomly generated MNFP with the characteristics of the LTL freight distribution problem. The underlying networks have the time-space structure necessary to model the dynamic nature of scheduling decisions and the commodities are small relative to the link capacity. Though the LTL networks generated and solved are too small to represent practical scheduling problems, they represent very large MNFP which were previously unsolvable. The problems of the second data set have a very different underlying structure and the commodities are large relative to the link capacity.

Two implementations of Dantzig-Wolfe decomposition are tested. In the first, a set of columns is added to the Master Problem in each iteration, the more common implementation. In the second, a single column is added in each iteration, the implementation which most resembles the PPLP. Our results indicate that PPLP solves the LTL freight distribution problem on average 20 times faster than the first implementation of Dantzig-Wolfe decomposition and on average, PPLP solves problems in the second data set about twice as fast. The PPLP algorithm solved the LTL routing problems on average 30 times faster and up to over 100 times faster than the LP solution with the default or all slack initial basis. When the links in the initial set of minimum path trees are specified in the initial basis, the LP solution is accelerated by a factor of up to 25 times for the LTL networks and from 2 to 4 times for the networks of the second data set. These results alone are interesting and indicate the great computational savings which can be realized if the structure inherent in the problem is exploited.

The major contributions of this dissertation are in numerical optimization, but it has also yielded new insights into Dantzig-Wolfe decomposition using *path* flows for each origin-destination pair instead of *tree* flows. In the course of developing this algorithm, our attention was drawn to the different problems that can be modeled as MNFP. A commodity may be flow between a single origin and a single destination, or flow from multiple origins into a single destination (or conversely, flow from a single origin into multiple destinations) or a commodity may have multiple origins and destinations.

There is a general perception that MNFP with large numbers of commodities and path-based formulations yield intractable problems. In fact, the algorithm developed here exploits this structure. We speculate that recasting a problem so that it can be modeled on path flows, either by decomposing tree flows into path flows or by adding super sources and sinks as required, reduces the number of extreme points enumerated. These alternative models also greatly increase the number of commodities, and thus the number of convexity (flow conservation) constraints of the Dantzig-Wolfe Master Problem, but this larger Master Problem can be solved by the specialized primal partitioning algorithm, PPLP.

This dissertation was completed at Princeton University under the supervision of Professor Warren Powell.

Warehouse Location with Service-Sensitive Demand

by

PENG-KUAN HO

The design of a distribution Network involves many interdependent decisions which can be classified into three components, facility, transportation, and inventory decisions. The decisions of each of these three components affect not only the cost of that component, but also the costs of the other components. Therefore, the design of a distribution network should consider the trade-off among the costs of all three components. In addition, there is also an interdependence between distribution network design and demand. The demand and its geographical distribution affect the optimal design of a distribution network, which in turn affects demand through its effect on the level of customer service. During the last two decades, there has been a growing emphasis on the customer service element of physical distribution, due to changes in the structure of our modern economy, as well as due to technological developments in the areas of information systems and communication. Consequently, it has become signficantly more important to explicitly consider customer service, in the design of a distribution system.

Among the most important distribution network design decisions are warehouse location decisions, since they have significant long term impacts on both distribution cost, and on the level of customer service provided by the distribution system. A significant amount of research has been devoted to the development of warehouse location models. However, most of these models do not consider the entire interrelationships between the three decision components of distribution network design. Furthermore, they do not take into account the interdependence between distribution network design and demand. The objective of this research is to develop a warehouse location methodology which considers the relationship between demand and customer service. Such a methodology should recognize the interdependence between warehouse location decisions, inventory decisions, customer service and demand. Therefore, our study starts with a comprehensive analysis of this interdependence. Based on the analysis, we define the Service-Sensitive Warehouse Location Problem (SSWLP) as that of determining the number, sizes, and locations of warehouses, the allocation of markets to established warehouses, the service level to be provided for each market, and the demand generated by each market, so as to maximize total profit.

The relationship between demand and customer service is represented by the "service-sensitive demand function." this function is developed in three steps, 1) selection of "service elements," 2) formulation of the "service-elements" as "service variables," 3) specification of a mathematical relationship between demand and the "service variables." Previous studies indicate that product availability and order cycle time are the two elements of customer service which most significantly affect demand. We select the "fraction of unsatisfied demand" and average order cycle time as the "service variables" to represent product availability and order cycle time, respectively. Two functional forms of demand function are considered, Linear and Constant Elasticity Demand Functions.

We formulate the SSWLP as a nonlinear mathematical programming model, where the objective function represents the maximization of total profit. Under the assumption of perfectly competitive market, price is considered given, while total cost is the sum of distribution costs and non-distribution costs, e.g., production cost, marketing cost, etc. The distribution costs consists of warehousing, transportation, and inventory costs. The SSWLP model includes three groups of constraints, 1) standard warehouse location constraints (common to existing models), 2) demand definition constraints, and 3) service requirements constraints. The SSWLP model is a large nonlinear model with both continuous and integer variables. Such a model cannot be solved using existing exact optimization techniques. We develop a solution method for the SSWLP model, which is based on decomposing the SSWLP into three subproblems, Warehouse Location Subproblem (WLSP), Warehouse Throughput Subproblem (WTSP), and Demand Generation Subproblem (DGSP).

Given the demand of each market when allocated to each of the potential warehouse locations, the WLSP determines the optimal number and locations of warehouses, and allocation of markets to warehouses. The WLSP is formulated as an Integer Programming Model and solved using Branch and Bound. To improve the computational efficiency in solving the WLSP, we present a reduction method based on the prespecified requirements on order cycle time. Additional reduction in the number of integer variables is achieved by adding integrality constraints only as they become necessary.

Given the established warehouses and allocation of markets to warehouses, the WTSP determines the total throughput at each of the established warehouses. The WTSP is formulated as a non-linear programming model with both integer and continuous variables. The WTSP can be decomposed into several independent subproblems, one for each established warehouse. The solution method for each of the WTSP's subproblems depends on whether the warehouse capacity constraints are binding or not. For the case of a nonbinding capacity constraint, we prove that the subproblem has a unique optimal solution that can be obtained using a simple numerical method. For the case of a binding capacity constraint, only a subset of the markets allocated to the warehouse can be served; for that case we propose a solution method which first determines the markets to be served by solving a "knapsack" problem, and then determines the throughput at the warehouse as in the case of a nonbinding capacity constraint.

Given the "expected" throughput of each potential warehouse location, the DGSP is the problem of determining the demand generated from each market when allocated to each of the feasible potential warehouse locations. The DGSP for a given warehouse-market pair can be solved by substituting directly the "expected" throughput at the warehouse and the order cycle time between the warehouse-market pair, in the service-sensitive demand function. The proposed solution process involves solving the three subproblems sequentially and iteratively, until either one of the following two conditions are satisfied: 1) the solutions of the WLSP at any two consecutive iterations remain unchanged, and 2) the prespecified maximum number of iterations is reached.

To test the validity of the SSWLP model and the proposed solution method, and to investigate the warehouse location behavior under service-sensitive demand, we conduct a comprehensive analysis in which we apply the model to three different test problems, with both capacitated and uncapacitated warehouses. The validity of the proposed SSWLP model and solution method is tested by comparing the SSWLP solutions with those obtained from optimally solving the location problem under inelastic demand. The results show that the proposed SSWLP model solutions provide significantly larger profits than those obtained from solving the warehouse location problem under inelastic demand. As expected, when demand becomes more sensitive to service, it becomes increasingly more important to explicitly represent the elastic nature of demand in location analysis. Unlike warehouse location analysis with inelastic demand where all the demand needs to be served, a warehouse location model with service-sensitive demand selects the profitable markets to be supplied.

We analyzed the effect of demand behavior on location behavior by comparing the model's results under the Linear and the Constant Elasticity Demand Functions. The results indicate that the number of established warehouses is greater under the Constant Elasticity Demand Function. The differences in location behavior under the two functions increase with the number of potential warehouse locations and potential markets. The presence of capacity limitations at the warehouses may change the effect of demand behavior on location behavior. A comprehensive sensitivity analysis shows the capability of the proposed model to provide valid and consistent responses to changes in various characteristics of the distribution system, as well as to changes in demand characteristics.

This disseration was completed at the University of Maryland under the supervision of Professor Jossef Perl.

Improving Railroad On-Time Performance: Models, Algorithms and Applications

by

DEJAN JOVANOVIĆ

This dissertation presents a methodological framework and tools that aim to improve the on-time reliability of railroad service. It represents one of the first attempts to conceptualize a hierarchy of connected models that support decision-making at various levels of the railroad organization, and to define the flow of information among these models. In this framework we concentrate on the purpose, scope, and mechanics of the decision support tools for tactical train scheduling and train dispatching rather than presenting a comprehensive and complete view of all aspects of railroad operations.

Poor, inconsistent on-time performance and the resulting low level of service is one of the most important problems of freight railroads in many developed countries; in developing countries, inadequate capacity of mostly single-tracked railway lines and scarce funds for capital improvements are becoming the major problems. Many railroads in North America are facing a combination of the abovementioned problems. Rising levels of traffic, predominantly single-tracked lines, and limited capital for physical track capacity improvements create increasing levels of congestion on the lines and in terminals, and make it very difficult for these railroads to provide consistent service to their customers. At the same time, American railroads are hard-pressed by their customers and by competition from the trucking industry to improve the consistency of shipment delivery times.

A seeming contradiction is that North American railroad traffic is increasing despite relatively poor service. This phenomena can be partially explained by the breakdown of traffic growth by commodity. The biggest increase has been in bulk goods traffic, such as grain, coal, and ore-commodities that are not very sensitive with regard to shipment delivery time. However, trains with not-so-sensitive lading share the same tracks with highly time-sensitive intermodal trains and mixed-merchandise trains and contribute to overall congestion. A segregation of trains by the level of service required by their lading that enables service level commitments for various commodity types to be met (i.e., level of service according to the price that the customer was charged) is essential in order to preserve customer satisfaction and railroad revenue. There is ample empirical evidence that the existing system of train priorities, in which train dispatchers, in resolving a conflict, always delay the lower priority train regardless whether trains are late or early or have a connection to make, is not sufficient for this purpose.

This dissertation addresses the above issues by presenting decision support tools aimed at the better capacity utilization and more consistent operations through optimal computer-aided train dispatching (CAD) in real-time.

OPTIMAL TRAIN DISPATCHING

THE BIGGEST technical obstacle to the successful implementation of an optimal CAD system is the combinatorial nature of the optimal train dispatching problem and the need for optimization algorithms that could provide good solutions in a rapidly changing, real-time environment. (There are also some institutional and managerial constraints as well, such as providing adequate authority and incentives to train dispatchers.) We develop novel, highly specialized, lower-bound based algorithms for the minimum tardiness cost train dispatching problem: an exact algorithm and a heuristic. It is shown by extensive numerical tests based on real-world data that the proposed algorithms represent a dramatic improvement upon the current state-of-the-art, enumeration-based methods.

The new algorithms are fast enough to enable a single CAD system to handle larger traffic volumes and cover longer planning horizons and larger dispatching territories then currently possible, while providing for optimal or near-optimal ways to move the trains over the railroad line. The extended time and distance scope of CAD systems based on the new algorithms facilitates coordination among various dispatching territories; it also minimizes the need to replan the dispatching plan due to "unexpected" train arrivals that could have been known, but were simply too far away in time or space to be included in the scope of the current dispatching plan. The algorithmic speed also provides the capability for extensive "what if"-analysis by the dispatcher.

The new minimum tardiness cost algorithms are especially well suited for dispatching problems where the goal is to meet target train arrival times, and to minimize the system cost of late arrivals when these are unavoidable, rather then to minimize total train delay as has been the usual objective of previous dispatching models. The algorithms can handle any nondecreasing tardiness cost function associated with a train, including nonlinear, nonconvex and nondifferentiable functions. In fact, the realities of railroad operations provide examples where the cost of train lateness is a staircase-shaped function. For example, a U.S. federal rule prohibits train crews from operating a train for more than 12 consecutive hours; if a train is not at a crew-change station within exactly 12 hours from the previous crew change, a high cost (a fine, or stopping the train and bringing a fresh crew-on line) is incurred. Another example of a nondifferentiable cost of late train arrival can be the cost of missing the connection at a yard for multiple blocks of cars with different connection times that are hauled within the same train; the train lateness cost 'steps-up' as the connection time for each of the blocks is missed.

The optimal train dispatching framework proposed in this dissertation attempts to fill the void in the literature regarding the use of and the benefits from a CAD system. After a review of the most commonly used CAD objective functions and costs affected by train dispatching, it is argued that the primary purpose of dispatching tools should be to allow trains to arrive *on-time* or on-target rather than to minimize total train delays. It is also argued that dispatchers should have more control as to when trains enter the line from the yards and at junctions. The windows for track maintenance gangs can be treated as special, very slow and very long trains within the CAD system in order to minimize their impact on trains. The potential of well-designed optimal computer aided dispatching systems to minimize train lateness is illustrated by examples.

There are some arguments within the railroad industry that, except for passenger and some intermodal trains that are not reclassified at yards, on-time train performance is not important; rather it is the reliable connections of cars transferring between trains in yards that determine on-time performance of rail shipments. We agree; adherence to train schedules (except for passenger and other highly sensitive trains) is not the goal in itself. However, the coordination of target train arrival/departure times (that may differ from published schedules) in real time to facilitate car connections, and realization of these target times through system-wide coordinated dispatching seems to be the only way to ensure reliable car connections. Thus, coordination of dispatching among different territories and setting of target train times over the whole railroad network in real time (real-time train scheduling) remains an open research issue.

TRAIN SCHEDULING DECISION SUPPORT

TACTICAL TRAIN scheduling is the planning aspect of train operations; only an achievable and robust tactical plan (schedule) can produce reliable and consistent real-time operations (train dispatching, train processing in yards) and result in a high level of rail shipment on-time performance.

A tactical train scheduling decision support methodology that uses the minimum tardiness train dispatching algorithms was implemented within and illustrated by the Schedule Analyzer (SCAN II) software package. The new methodology offers a different approach from simulation models traditionally used to analyze railroad operations. The SCAN methodology supports train scheduling by answering the following questions:

- Given a set of train schedules and physical line and train characteristics, is there a way to achieve the given set of schedules, i.e., are they feasible?
- If a given set of schedules is not feasible, what is the feasible set of schedules closest to the given one?
- If a given set of schedules is feasible, what is the expected on-time performance when stochastic operating elements are introduced?

SCAN focuses on schedules by removing bias introduced by fixed train priorities and historical and/or stochastic data commonly used by simulation models.

The use of SCAN by a large North American railroad and evaluation of a number of real-world train schedules has confirmed the hypothesis that the poor on-time performance is in part due to the inadequate scheduling. However, feasibility of a given set of schedules does not ensure that they will be consistently operated in the relatively stochastic environment of North American railroads. First, a significant proportion of trains, mainly grain and coal unit trains, are not scheduled at the tactical level (however, as SCAN methodology proposes, time slots for unscheduled trains can be scheduled at the tactical level). Second, changes in traffic demand may result in addition of extra, unscheduled "sections" of scheduled trains; subject to locomotive and crew availability and yard processing times trains can depart late, or even early from terminals. Third, mechanical breakdowns, bad weather, and derailments add additional uncertainty to railroad operations. Stochastic elements can be introduced into SCAN to estimate on-time performance, or to perform a *what-if* analysis, similar to the use of simulation models but with a focus on schedules. However, the creation of *robust* or reliable train schedules remains a challenge.

This research opens as many new questions as it resolves. Suggestions and guidelines for future research and extensions of the developed methodology are presented at the end of each chapter.

This dissertation was completed at the University of Pennsylvania under the supervision of Professor Partrick Harker.

Equilibrium Behavior of Markovian Network Processes

by

KWANGHO KOOK

How long does it take to travel from one location to another in a Jackson queueing network? This simply stated problem had challenged researchers for over two decades. The main contribution of this dissertation is a solution to the problem. Earlier studies had shown that, for a Jackson network in which each node is a single server and units do not overtake one another as they move among the nodes, the travel time through several nodes is a sum of the independent exponentially distributed sojourn times at the nodes. Unfortunately, this result is not true for networks in which a node may have several servers or units may overtake one another. The difficulty is that, without these restrictive assumptions, the sojourn times at the nodes are generally not independent and not exponentially distributed. Consequently, travel times are rather complex.

Our study considers a broader issue of characterizing travel times for a large class of "routes" in Markovian networks. The networks are generalizations of Jackson networks; and the nodes may have several servers or processor sharing, and units may overtake one another. We introduce a general notion of a "route" that a unit may take in the network; the "travel time" on the route is the time a unit takes to traverse the route when the network is stationary or in equilibrium. Examples include the time to travel from one sector of the network to another, the time a unit spends in a certain sector, the time between visits to a certain sector by a unit and the time that a unit carries a certain label denoting its type (which is changeable) during its sojourn in the network. Our main results are expressions for the means and, in some cases, higher moments and Laplace transforms of these travel times. The difficulty in analyzing these times is that it is not known whether a unit is traversing a route until the route is completed; a unit may start on a route and not complete it. For example, a unit beginning to travel from one sector to another may return to the first sector to begin again. Consequently, the number of units undergoing a passage on a route at any time is a function of the future of the network process as well as its past. To overcome this difficulty, we have devised a novel device for labeling the units that allows us to look into the future of the network process; this monitors those units that do indeed complete the route. Using this and certain laws of large numbers for regenerative processes, we derive expressions for mean travel times. For certain routes whose input flow is Poisson, we obtain Laplace transforms and moments of travel times.

Although the rest of this dissertation appears to be less exciting than the one described above, it may be just as important in the long run, especially for the development of intelligent networks. The general theme of our study is to characterize the equilibrium behavior of networks in which discrete units move among nodes where they are processed. Some archetypal networks are as follows:

Flexible Manufacturing Networks: Parts, tools, pallets, AGVs or material (units) move among a group of work stations and storage areas (nodes) that machine and store the units for later use or for shipping.

Computer Networks: Transactions, tasks, queries, data packets or programs move among processors, computers, peripheral equipment or files.

Telecommunications Networks: Telephone calls, data packets or messages move among operators or switching stations.

Maintenance and Logistics Networks: Repairable parts or equipment needed for the operation of a large system move among locations where they are used, repaired, and stored.

Distribution Networks: Goods, orders or trucks move among plants, warehouses or market locations.

Biological and Chemical Networks: Animals, cells, molecules, neurons, etc. move among locations or change their states or shapes.

There is an extensive theory of Jackson, MCMP and Kelly networks in which the nodes operate independently and units move one at a time and their routes are independent. The equilibrium distributions of these networks are of product form and are well understood.

By its very nature, however, a network is a system of interacting nodes: the processing at the nodes and the routings typically depend dynamically on the actual congestion, and units move concurrently. Examples of dependencies are routing of units to avoid congested nodes, speeding up of processing as queues grow, splitting and merging of units, and batch or distributed processing. Very little is known about such networks. A major hindrance has been the lack of intuition as to what type of multivariate probability distribution might be appropriate for the equilibrium distribution of the locations of units in the network. We have found such a distribution, and have used it to characterize a large class of networks with dependencies.

Our approach is based on the observation that all the networks with tractable equilibrium distributions satisfy a certain "partial balance" property. Noting that this property was relevant to Markov processes in general, we have formulated an abstract Markov process with bivariate transition operators that can be used as a vehicle for studying networks. We derive a set of necessary and sufficient conditions for this Markov process to be partially balanced, and give an expression for the equilibrium distribution of a partially balanced process. We use these results to describe the equilibrium distributions of several general families of networks with dependencies and concurrent movements as well. We present particular distributions for tree-like networks, clustering processes, and logistics networks with resource sharing. Our study of concurrent movements also led us to equilibrium distributions for certain batch arrival and batch service queues; these are for single service systems whose behavior we need to understand in order to analyze networks of such systems.

The results in this second part of our work can be viewed as a nucleus of a theory of partially balanced network processes. Such knowledge is needed to understand the performance of the next generation of intelligent networks that will be the backbone of our computer, communications and transportation systems.

This dissertation was completed at the Georgia Institute of Technology under the supervision of Professor Richard Serfozo

Warehouse Location under Multiple Transportation Options

by

SOMPONG SIRISOPONSILP

Logistics decisions and logistics costs can be classified into three inter-related components, facility location, transportation, and inventory. Each of these components of logistics decisions affects not only the cost of that component, but also the costs of the other two components. The regulatory reform of the U.S. freight transportation industry has increased significantly the spectrum of transportation options available to shippers, thereby increasing the importance of representing multiple transportation options and the interdependence between the above three components, in the design and analysis of a logistics system.

Among the most important problems in logistics management that need to be analyzed under multiple transportation options is that of warehouse location. Given the interdependence between location, transportation, and inventory decisions, location analysis under multiple transportation options should consider the selection of transportation options as a design element and should explicitly represent inventory costs. Existing warehouse location models commonly treat transportation choice as input, rather than output. Furthermore, they typically do not include an explicit representation of inventory costs. *The objective of this study is to develop a methodology for analyzing warehouse location under multiple transportation options, which recognizes the interdependence between location, transportation, and inventory decisions.*

The Combined Warehouse Location-Transportation Problem (CWLTP) is that of determining the number, sizes and locations of warehouses, and the "optimal" transportation options between plants and warehouses. We formulate the CWLTP as a nonlinear programming problem with both integer and continuous variables. The proposed CWLTP model differs from existing warehouse location models in three important aspects. First, it considers the selection of transportation options as output rather than input. Second, it includes an explicit representation of the inventory implications of warehouse location and transportation decisions. Third, it explicitly represents the required level of customer service.

The CWLTP is a large and difficult problem which cannot be solved directly using existing exact optimization techniques. The proposed method for solving the CWLTP is based on decomposing the CWLTP into two subproblems, Warehouse Location-Allocation Problem (WLAP) and Transportation Option Selection Problem (TOSP). For a prespecified set of selected transportation options between plants and warehouses, the WLAP optimizes the locations of warehouses and the allocation of demand points to warehouses. Given the locations of the established warehouses and the throughput at each warehouse as determined from the solution to the WLAP, the TOSP selects the "optimal" transportation option for each potential plant-warehouse pair.

Given the selected transportation options between plants and potential warehouse locations, we show that the CWLTP reduces to the WLAP. Like the CWLTP, the WLAP is a difficult nonlinear programming problem with both integer and continuous variables. The proposed strategy for solving the WLAP consists of, 1) simplifying the WLAP and solving the simplified WLAP exactly, and 2) adjusting the solution for the simplified WLAP toward an improved solution for the original WLAP. The simplification of the WLAP is based on the criterion of minimizing the sum of "overshoot" and "undershoot" caused by the simplification. The simplified WLAP is a linear mixed integer program; it is solved optimally using the Functional Mathematical Programming System (FMPS) on the UNISYS 1100 system. Given the locations of warehouses and the assignment of plants to warehouses, the proposed adjustment procedure, termed the Demand Reallocation Procedure, attempts to improve the allocation of demand points to warehouses.

In the context of the TOSP, we distinguish between "active" and "inactive" trunking lanes. "Active" trunking lanes refer to the trunking lanes which form the distribution network, while "inactive" trunking lanes are those excluded from the network. Solving the TOSP for "active" and "inactive" trunking lanes requires different approaches based on different assumptions. Given the WLAP solution, the TOSP for "active" trunking lanes can be derived directly from the CWLTP formulation under the assumption that the flow over any trunking lane remains unchanged. The resulting TOSP for "active" trunking lanes is an integer programming problem with a nonlinear objective function. We show that the TOSP model for "active" trunking lanes can be decomposed into a set of independent Transportation Option Selection Subproblem (TOSS), one for each open warehouse. We develop an implicit enumeration scheme for solving the TOSS exactly.

"Inactive" trunking lanes can be further classified into those leading to open warehouses, and those leading to closed warehouses. In the case of an "inactive" trunking lane leading to an open warehouse, the total throughput at the warehouse is known and all the cost components can be estimated based on that throughput. For this case, the TOSP reduces to the TOSS, under the assumption that the selection of transportation options in that trunking lane is made as if the warehouse would be supplied entirely via that lane.

In the case of an "inactive" trunking lane leading to a closed warehouse, both the flow to be shipped via that trunking lane and the throughput of the warehouse associated with the lane, are not known. In this case, we develop an approach to estimate the expected throughput of the closed warehouse, and subsequently show that given the "estimated" throughput the TOSP reduces to the TOSS.

To test the validity of the solutions provided by the proposed CWLTP algorithm and its capability of solving problems of reasonable size, we apply the algorithm to a sample of test problems. We also conduct a sensitivity analysis to investigate the capabilities of the model to provide reasonable and consistent responses to changes in operating policy, and in the environment in which the distribution system operates.

The analysis shows that the CWLTP solution algorithm can solve relatively large problems with 3 plants, 24 potential warehouse locations, 50 demand points, and 3 transportation options, but with substantial computational effort. For each test problem, the percentage breakdown of total distribution cost into major cost components is consistent with those reported by previous studies. The application of the proposed CWLTP algorithm to problems of different sizes under "homogeneous" conditions clearly shows the interdependence between transportation decisions and warehouse location decisions. The results show that when the set of available transportation options between every plant-warehouse pair is increased, a significantly larger reduction in total distribution cost can be achieved by optimizing simultaneously the locations of warehouses and the choice of transportation options relative to the sequential approach of optimizing transportation decisions after location decisions.

The sensitivity analysis focuses on new elements of the CWLTP model, which are not represented explicitly in existing warehouse location models. The results of the sensitivity analysis show that the proposed CWLTP algorithm provides consistent and reasonable responses to changes in the availability and characteristics of transportation options, unit inventory cost, and the prespecified level of customer service. While the number of warehouses is shown to be relatively insensitive to these changes, the locations of warehouses may be significantly affected by both the availability and the quality of transportation services, as well as by a change in unit inventory cost. The number and locations of warehouses are shown to be relatively insensitive to the company's policy regarding product availability.

The contributions of this study can be summarized as follows: 1) improved understanding of the interdependence between location, transportation, and inventory decisions, 2) definition of the CWLTP, 3) mathematical programming model for the CWLTP, and 4) heuristic algorithm for the CWLTP model. Most importantly, the proposed methodology for warehouse location under multiple transportation options provides a more accurate representation of the warehouse location problem in a deregulated environment where the spectrum of transportation services is increased.

This dissertation was completed at the University of Maryland under the supervision of Professor Jossef Perl.

Period and Phase of Customer Replenishment: A New Approach to Inventory/Routing Problems

by

IAN R. WEBB

This dissertation develops effective new approaches to both the strategic and tactical inventory routing problems. Much of the research is based on the use of period and phase of customer replenishment as decision variables. These decision variables have not previously been associated with vehicle routing problems.

Inventory routing problems (IRPs) arise in distribution settings when the model to be used must include both inventory management and vehicle routing characteristics if it is to adequately represent the system of interest. Important potential applications for such models include fuel and gas distribution, garbage collection and the supply of retail outlets. IRPs can be divided into two classes: tactical and strategic. Tactical inventory routing models are concerned with minimizing the sum of transportation and inventory costs incurred in the day to day operation of a distribution system, subject to constraints on the number and capacity of available vehicles, maximum route length and/or duration, vehicle/customer compatibility and customer storage capacity with respect to customer product consumption rates. In such models it is generally assumed that the statement of the problem includes the location, demand characteristics and storage capacity of the customers and the composition and carrying characteristics of the vehicle fleet. Strategic inventory routing models on the other hand are more concerned with the design of the distribution system itself, particularly the size and composition of the vehicle fleet to be used, given some initial information regarding the customers to be served. The focus is typically the minimization of the total cost of acquiring the vehicle fleet or construction of other necessary system components.

The Strategic Inventory Routing Problem (SIRP) is the first of the first class of problems we study. Given the location, mean and standard deviation of daily demand, storage capacity and maximum acceptable probability of stockout for each set of customers, and the carrying capacity and travel speed of the vehicles to be used, the objective in SIRP is to determine the minimal size vehicle fleet required to supply customers with "product" (e.g., industrial gasses, empty space) from a central depot. This work builds on recent successful work by LARSON[2] in New York City by generalizing the basic SIRP model developed for that application. The SIRP model has a wide range of potential applications in the design of systems and is not necessarily restricted to classical inventory routing problems. The "vehicle fleet" might be any form of server and the inventory may be interpreted as service of some kind rather than simply as a physical product. Larson's approach involves an initial transformation of the given probabilistic demand problem into an equivalent problem with deterministic customer inventory consumption. Customer clusters and associated replenishment routes are then developed for this deterministic problem. The vehicle requirements for each cluster are calculated and used in determining the overall fleet estimate. Since even the deterministic problem is NP-hard, the solution approach developed is a greedy heuristic based on the Clarke-Wright "savings" approach.

This dissertation begins by studying the original SIRP model in detail and shows that the vehicle requirements for any given cluster can be expressed as a function of customer and vehicle parameters. These expressions lead to several conclusions regarding the character of optimal solutions for SIRP. This work also leads to the development of two mathematical programming formulations for SIRP—one a transshipment model and the other a set covering model. Both formulations involve complex nonlinear objective functions. Possible alternative solution approaches suggested by these formulations are discussed.

A unique characteristic of the basic SIRP model in [2] was that the intervals between replenishments of customer clusters are determined as part of the solution. In previous work on inventory routing type problems involving "periodicity," the intervals between customer replenishments have been pre-specified in the statement of the problem, e.g., [1, 4, 5]. The success of the SIRP model is largely due to this additional flexibility. We generalize the basic SIRP model by employing period and phase of *customer* replenishment as decision variables. The period and phase of a customer determines the timing of that customer's replenishments relative to those of other customers within the cluster involved. This approach leads to *customer*-specific replenishment requirements. Efficient routes are designed to meet these requirements and in turn lead to improved strategic decisions on capital investments and manpower planning. The generalized SIRP model is quite complex and useful mathematical programming formulations were not found. Two heuristics based on Larson's original solution procedure were developed. Each of the procedures uses a greedy search heuristic to determine the period and phase of replenishment for individual customers.

The period/phase approach also leads to a new approach for the Tactical Inventory Routing Problem (TIRP). The version of TIRP studied involves N_c spatially dispersed customers. The location of each customer is known, as is the actual inventory level at the start of each day and the probability distribution function of each customer's daily inventory consumption. Inventory is delivered by a fleet of N_v identical vehicles, with each vehicle having a carrying capacity of C_{veh}, travel speed V and a maximum shift length L. Vehicles leave the depot at the start of the day and return at the end of their shift. Shifts are less than 24 hours in length, hence all vehicles are available at the start of each day. At the start of each day the vehicle dispatcher must:

(i) decide which customers to replenish and
(ii) develop an efficient replenishment scheme

subject to the constraints imposed by the vehicle fleet size. The objective is to minimize total system cost, which includes the future costs of current decisions. The uncertainty of future demand at customers makes these future costs uncertain. Suppose we formulate a Markovian decision problem in which the states correspond to possible states of the system wide inventory levels and the decision alternatives involve the feasible replenishment schemes. Solving this MDP would yield the expected future cost of being in any state at the end of the current day. By weighting the expected future costs of the states by the appropriate transition probabilities, the expected future cost (and hence the total expected cost) of a given action can be calculated. In the past, the major obstacles in applying such a Markovian decision-theoretic approach to inventory routing problems were that the size of the state space required is generally prohibitively large for all but relatively small problems and that the number of possible replenishment schemes is combinatorially large. Our approach substantially overcomes both these obstacles.

In the course of developing a fleet estimate for SIRP, a set of "model" replenishment routes and clusters are developed as functions of relevant customer, vehicle and geographical characteristics. Hence, it is reasonable to expect that an optimal or near-optimal solution to the associated daily instances of TIRP can be constructed from routes identical or similar to those generated in arriving at the SIRP solution. A heuristic for TIRP which restricts the search to those routes generated in the solution to SIRP should generate solutions of relatively good quality. Further, since each of these routes visits customers from only one SIRP cluster, the use of these routes as the available strategy set results in a natural decomposition of TIRP into a number of inventory routing subproblems, each corresponding to one of the clusters in the SIRP solution. These cluster subproblems are linked by vehicle availability constraints. Based on this reasoning, our solution approach involves two steps:

(i) solving the Markovian decision subproblem for each cluster to determine total expected costs for each of the candidate routes in that cluster, and
(ii) solving a coordination problem to determine which routes will actually be performed in order to minimize total expected future cost subject to the available fleet capacity.

Each of the subproblems is much smaller than the original instance of TIRP. The Markovian state space is small enough that the Markovian decision problem for each cluster can often be solved exactly. For instances where these subproblems are too large for exact solution, an approach based on Minkoff's Future Value Decomposition heuristic[3] is developed. The coordination problem linking the cluster subproblems takes the form of a new and complex knapsack-like problem. A heuristic solution procedure is developed for this problem.

The results of computational tests of the algorithms developed for both SIRP and TIRP are included. Problems with up to seventy customers were solved on a Macintosh personal computer. The generalized SIRP model yields fleet size estimates equal to or lower than those from the original model. The observed reductions in the fleet size estimate are typically one or two vehicles. Such reductions can result in substantial savings in both capital and operating costs. The Markovian decision-theoretic approach for TIRP appears to be very effective in terms of minimizing stockouts while keeping customer inventory levels low, consistent with modern Just-In-Time approaches.

A number of interesting and challenging research questions are identified in this dissertation. One area of future research will be the relaxation of the more restrictive assumptions on the information available. As an example, it is currently assumed that the exact inventory levels at each customer are known. In many applications, it is more likely that only estimates of the inventory levels will be available. Other work will focus on multiple vehicle types and inventory categories, improved solution algorithms and the development of lower bounds on optimal solutions based on the mathematical programming formulations of the SIRP models.

REFERENCES

1. N. Christofides and J. Beasley, "The Period Routing Problem," *Networks* **14,** 237-256 (1984).
2. R. Larson, "Transporting sludge to the 106-Mile Site: An Inventory/Routing Model for Fleet Sizing and Logistics System Design," *Trans. Sci.* **22,** 186-198 (1988).
3. A. Minkoff, "Real-time Dispatching of Delivery Vehicles," Ph.D. thesis in Operations Research, Massachusetts Institute of Technology, Cambridge, Mass., 1985
4. R. Russell and W. Igo, (1979): "An Assignment Routing Problem," *Networks* **9,** 1-17 (1979).
5. C. Tan and J. Beasley, "A Heuristic Algorithm for the Period Vehicle Routing Problem," *Omega* **12,** 487-504 (1984).

This dissertation was completed at the Massachusetts Institute of Technology under the direction of Professor Richard Larson.

Published in Transportation Science on August 01, 1991 as DOI: 10.1287/trsc.25.3.245.
This article has not been copyedited or formatted. The final version may differ from this version.