TRISTAN 1

TRIENNIAL SYMPOSIUM ON TRANSPORTATION ANALYSIS

SYMPOSIUM INTERNATIONAL SUR LA RECHERCHE EN TRANSPORT

MONTRÉAL - QUÉBEC

JUNE 5 - 11, 1991

Program and Abstracts

Programme et résumés

Centre de recherche sur les transports
Université de Montréal
C.P. 6128, succursale A,
Montréal Canada H3C 3J7

Telephone: (514) 343-7143
Telex: 05-24146 BIBPOLYTEC
Email: TRISTAN@crt.umontreal.ca
Fax: (514) 343-7121

Dear colleagues and friends,

It is our pleasure to welcome you to the first TRISTAN Conference. We believe that the meeting will prove scientifically exciting and rewarding, as well as socially enjoyable.

TRISTAN was born of the realisation that, while research in transportation and logistics has matured significantly over the last twenty years, a forum was needed to target the needs of this area, with a strong emphasis on identifying and advancing important areas of research. One the main goals of TRISTAN is to bring together in a collegial research atmosphere the people actively involved in fundamental and applied research directly related or indirectly relevant to transportation.

The format of the conference has been designed to focus attention on the most important research results, to allow the broadest possible involvement of all participants in all activities, and to foster the exchange of ideas. The program of TRISTAN is a reflection of the richness and diversity of our field. It includes theoretical and algorithmic results, as well as good experimental work and creative applications, since it is felt that healthy progress at a theoretical level requires a strong presence from the experimental and applications community. It is our expectation that TRISTAN will become the premier outlet for research in transportation science. We are extremely pleased that you are helping us achieve this goal by using TRISTAN as the forum for presenting your major results.

The goals set for TRISTAN are ambitious but due to the enthusiastic response of the community and with the assistance of a large number of people, we believe that they are being fulfilled. In particular, we would like to express our gratitude toward the members of TRISTAN's international advisory committee, who helped steer the conference and contributed their efforts to the reviewing process. We would also like to thank all the institutions and agencies which have contributed financially and have thus greatly facilitated the organization of TRISTAN. A special mention goes to the collegues and friends who were there at the very beginning: Mark Daskin, Pierre Dejax, Martine Labbé, and Marius Solomon. Finally, we want to thank particularly the local organizing committee, the Centre de recherche sur les transports and its staff, and all those who have contributed to make TRISTAN a reality.

TRISTAN complements the other general meetings and specialized conferences which cater to the needs of our community. We certainly do not wish to compete with these conferences but rather hope that a new tradition is being established, and that TRISTAN will continue to provide a high quality outlet for research related to transportation.

Warren B. Powell   Teodor G. Crainic

Hosted by:
Organisé par:

Centre de recherche sur les transports (C.R.T.)

Sponsored by:
Parrainé par:

AD OPT

Conseil de recherches en sciences naturelles et en génie du Canada / Natural Sciences and Engineering Research Council of Canada

École des Hautes Études Commerciales

École Polytechnique de Montréal

GIRO Inc.

Les Consultants INRO Consultants Inc.

Ministère des Transports du Québec

National Science Foundation (U.S.A.)

The Railway Association of Canada

Université de Montréal

Conference Chairmen
Co-responsables du Congrès

Teodor Gabriel Crainic
Centre de recherche sur les transports (C.R.T.) and D.S.A./UQAM

Warren B. Powell
Princeton University

Advisory Committee
Comité scientifique

Michael O. Ball
University of Maryland

Alex Belenky
USSR Academy of Science

Carlos Daganzo
University of California at Berkeley

Mark S. Daskin
Northwestern University

Jacques Desrosiers
Centre de recherche sur les transports (C.R.T.) and GERAD

Awi Federgruen
Columbia University

J. Enrique Fernandez
Catholic University of Chile

Marshall Fisher
University of Pennsylvania

Terry L. Friesz
George Mason University

Giorgio Gallo
University of Pisa

Patrick T. Harker
University of Pennsylvania

Alexander H.G. Rinnooy Kan
Erasmus University

Martine Labbé
Erasmus University

Gilbert Laporte
Centre de recherche sur les transports (C.R.T.) and GERAD

Jan Karel Lenstra
Eindhoven University of Technology

Thomas L. Magnanti
Massachusetts Institute of Technology

Amedeo R. Odoni
Massachusetts Institute of Technology

H. Donald Ratliff
Georgia Institute of Technology

Werner Rothengatter
Universität Karlsruhe

Jean-Marc Rousseau
Centre de recherche sur les transports (C.R.T.)

François Soumis
Centre de recherche sur les transports (C.R.T.) and GERAD

Local Organizing Committee (C.R.T.)
Comité d'organisation (C.R.T.)

Teodor Gabriel Crainic
Martin Desrochers

Michel Gendreau
Carole Laflamme

Gilbert Laporte
Clairette Simard

PROGRAM
PROGRAMME

Thursday, June 6, 1991 / Jeudi, 6 juin 1991

08:30-09:00 H.E.C. 3073	Opening session / Ouverture du congrès
09:00-10:30 H.E.C 3073	***Airline Crew Scheduling - Horaires d'équipages dans l'aviation civile*** "Optimizing the Airlines' Set-Partitioning Problems", K.L. Hoffman and M. Padberg "Column Generation Approaches to the Airline Crew Scheduling Problems", M. Desrochers, Y. Dumas, F. Soumis and P. Trudeau "An Alternative Solution Procedure for the Long-Haul Crew Assignment Problem", C. Barnhart, E. Johnson and R. Anbil
10:30-11:00	Break / Pause-café
11:00-12:30 H.E.C. 3073	***Stochastic Models I - Modèles stochastiques I*** "Routing with Stochastic Demands: Past, Present, and Future", M. Dror "Exact Solution of Some Transportation Problems Under Uncertainty", G. Laporte and F. Louveaux "Stochastic Programming Applications in Transportation: Dynamic Network Models of Common Carrier Operations", W.B. Powell, R.K.L. Cheung and L.F. Frantzeskakis
12:30-14:00	Lunch / Déjeuner (Université de Montréal)
14:00-15:00 H.E.C. 3073	***Focus 1*** "New Technology for Transportation", D.K. Willis
15:00-15:15	Break / Pause-café

15:15-16:30 H.E.C. 3072	***Equilibrium Models - Modèles d'équilibre*** (Parallel session / Session parallèle) "An Interactive Spatial Price Equilibrium Model for Designing Long-Run Strategies for Competitive Firms", A.E. de Rezende "Supply Demand Equilibrium on Transportation Networks with Composite Modes", E. Fernandez, J. de Cea and M. Florian "Modeling and Solving the Dynamic Traffic Assignment Problem with Windows of Time", O. Drissi-Kaïtouni, A. El-Merzouqi and M. Florian "A Conjugate Gradient Algorithm for the Dynamic User Optimal Traffic Assignment Problem", B.W. Wie
H.E.C. 3074	***Location and Networks - Modèles de localisation*** (Parallel session / Session parallèle) "Dynamic Facility Location Modelling Using Forecast Horizon Concepts: Issues, Algorithms and Results", M.S. Daskin and W.J. Hopp "On Locating Path- or Tree-Shaped Facilities on Networks", S.L. Hakimi, E.F. Schmeichel and M. Labbé "A Tabu Search Heuristic for Location-Allocation Problems with Balancing Requirements", M. Gendreau, T.G. Crainic, P. Soriano and M. Toulouse
16:30-16:45	Break / Pause-café
16:45-17:45 H.E.C. 3072	***Freight Transportation Models - Modèles pour le transport des marchandises*** (Parallel session / Session parallèle) "A Model, Algorithms and Strategy for Train Pathing and Planning", M. Carey "The Intermodal Trailer Assignment Problem: Models, Algorithms, and Heuristics", T.A. Feo, J.L. Gonzalez-Velarde
H.E.C. 3074	***Routing and Scheduling I - Routes et horaires I*** (Parallel session / Session parallèle) "A Two-Level Model for Helicopter Routing and Scheduling Solved by Column Generation", J. Desrosiers, Y. Dumas, M.M. Solomon and A. Chalifour "A Matching Based System for Solving a Special Class of Routing Problems", U. Derigs and A. Metz
18:00 H.E.C.	Wine & Cheese Party / Réception vin et fromage Offered by / Offert par: Les Consultants INRO Consultants Inc. (Salon des étudiants)

Friday, June 7, 1991 / Vendredi, 7 juin 1991

08:30-09:00 H.E.C. 3073	Mark Daskin *Transportation Science* Editor / Rédacteur
09:00-10:30 H.E.C. 3073	***Transit Network Models - Modèles de réseaux de transport en commun*** "A Variational Inequality Model for Transit Equilibrium Assignment", J.H. Wu, M. Florian and P. Marcotte "A Method for Optimizing the Frequencies in a Transit Network: A Special Case of Nonlinear Bi-Level Programming", I. Constantin and M. Florian "Route Generation and Improvement Algorithms for the Transit Network Design Problem", M.H. Baaj and H.S. Mahmassani
10:30-11:00	Break / Pause-café
11:00-12:30 H.E.C. 3073	***Exact Methods for the VRP - Méthodes exactes de résolution pour le VRP*** "A Cutting Plane Procedure for the Symmetric Generalized Traveling Salesman Problem", C.E. Noon and M.M. Sepehri "Capacitated Trees, Capacitated Routing, and Associated Polyhedra", J.R. Araque, L.A. Hall, T.L. Magnanti "Sensitivity Analysis Methods for Vehicle Routing and Scheduling Problems", P.M. Thompson and J.B. Orlin
12:30-14:00	Lunch / Déjeuner (Université de Montréal)
14:00-15:00 H.E.C. 3073	***Panel*** "Panel on Software Development and Implementation", L. Bodin, R. Dial, M. Florian, H.D. Ratliff and J.M. Rousseau
15:00-15:15	Break / Pause-café
15:15-16:30 H.E.C. 3072	***Stochastic Models II - Modèles stochastiques II*** (Parallel session / Session parallèle) "A Stochastic Programming Approach to O-D Matrix Estimation", S.W. Wallace and K. Jörnsten "A Stochastic, Dynamic Airline Network Equilibrium Model", F. Soumis and A. Nagurney "Reliability in Urban Transportation", B. Sansó and F. Soumis

H.E.C. 3074 ***VRP with Time Windows - Routes avec fenêtres de temps*** (Parallel session / Session parallèle)
"A Variable Depth Approach for the Single-Vehicle Pickup and Delivery Problem with Time Windows", L.L.J. Van der Bruggen, J.K. Lenstra and P.C. Schuur
"Vehicle Routing with Time Windows: Identification of Problem Structure", K. Halse
"A Two-Phase Heuristic for Vehicle Problems with Time Window Constraints", M. Haouari and P. Dejax

16:30-16:45 Break / Pause-café

16:45-18:00
H.E.C. 3072 ***Dynamic Shortest Paths - Plus courts chemins dynamiques*** (Parallel session / Session parallèle)
"Minimum Time Paths in Dynamic Networks with Application to Intelligent Vehicle/Highway Systems", D.E. Kaufman and R.L. Smith
"Dynamic Shortest Paths with Markovian Arc Costs", H.N. Psaraftis and J.N. Tsitsiklis
"Does Providing Information to Drivers Reduce Traffic Congestion?", R. Arnott, A. de Palma and R. Lindsey

H.E.C. 3074 ***Routing and Scheduling II - Routes et horaires II*** (Parallel session / Session parallèle)
"Heuristic Algorithms for the Multiple Depot Vehicle Scheduling Problem", M. Dell'Amico, M. Fischetti and P. Toth
"Expert Systems and Operations Research: Mutual Benefits for the Routing and Scheduling of Transportation Vehicles", J.Y. Potvin and J.M. Rousseau
"Arc Routing and Scheduling: New Algorithms and Applications", M. Ball, L. Bodin, H. Gun, L. Levy and I. Su

Saturday, June 8, 1991 / Samedi, 8 juin 1991

08:30-10:15	***Computational Methods - Méthodes numériques***
H.E.C. 3073	"Parallel Computation of Large-Scale Dynamic Market Network Equilibria via Time Period Decomposition", A. Nagurney and D.S. Kim "Solution of the Shortest Path Problem on Massively Parallel Computers", M.B. Habbal, H.N. Koutsopoulos and S.R. Lerman "A Splitting Equilibration Algorithm for the Computation of Large-Scale Constrained Matrix Problems: Theoretical Analysis and Applications", A. Nagurney and A. Eydeland "The Status of Massively Parallel Algorithms for Network Structured Problems", S.A. Zenios
10:15-10:45	Break / Pause-café
10:45-12:30	***Dynamic Traffic Assignment - Affectation dynamique du trafic***
H.E.C. 3073	"Algorithms for the Dynamic Network Equilibrium Problem", T.L. Friesz and R.L. Tobin "Solving the Dynamic User-Optimal Traffic Assignment Problem Over Congested Transportation Networks", D.E. Boyce, B. Ran and L.J. LeBlanc "Stochastic Process Modelling of Transportation Network Dynamics", E. Cascetta, G.E. Cantarella "An Algorithm for Dynamic Traffic Assignment Models Based on an Optimal Control Approach", J. Barceló and E. Codina
12:30-14:00	Lunch, then bus to Quebec City / Déjeuner, ensuite départ pour Québec

Sunday, June 9, 1991 / Dimanche, 9 juin 1991

Brunch (Château Frontenac, Salon Bellevue)

Monday, June 10, 1991 / Lundi, 10 juin 1991

09:00-10:30 Frontenac	***VRP Heuristics - Méthodes heuristiques pour le VRP*** "A TSSP+1 Decomposition Approach for the Capacity-Constrained VRP", J. Mittenthal and C. Noon "Solving Real-Life Vehicle Routing Problems with Tabu Search: Two Adaptations", F. Semet "A Tabu Search Heuristic for Vehicle Routing Problems", A. Hertz, M. Gendreau and G. Laporte
10:30-11:00	Break / Pausé-café
11:00-12:30 Frontenac	***Logistics - Logistique*** "The Location-Allocation-Inventory-Routing Model in the Design of Strategic Distribution Systems", M. Goetschalckx, S. Song and M. Cole "Continuous Modeling in Logistics", A. Langevin "Planning Models for the Distribution and Transportation of Containers", T.G. Crainic, M. Gendreau and P. Dejax
12:30-14:00	Lunch / Déjeuner
14:00-15:00 Frontenac	***Focus 2*** "Interior Point Methods for Linear Programming: Computational Results", I. Lustig
15:00-15:15	Break / Pause-café
15:15-16:15 Frontenac	***Air Transportation - Transport aérien*** "Dynamic Ground-Holding Strategies for Air Traffic Control", F. Somis and A.R. Odoni "Airline Seat Allocation with Multiple Nested Fare Classes", S.L. Brumelle and J.I. McGill
16:15-16:30	Break / Pause-café

16:30-17:45 Frontenac

Operations - Exploitation

"A Hierarchical Solution for the Service Network Design Problem", J.M. Farvolden

"An Analytical Model for Optimal Design of an Automated Guided Vehicle System", M.E. Johnson, M.L. Brandeau

"Using Game-Theoretic Methods in Decision Support Systems for Traffic Scheduling", A.S. Belenky and M.S. Dubson

19:00 Banquet (Restaurant Café de la Paix)

Tuesday, June 11, 1991 / Mardi, 11 juin 1991

09:00-10:15 Frontenac

Probabilistic Analysis for VRP Algorithms - Analyse probabiliste des algorithmes pour le VRP

"Probabilistic Analysis of Algorithms for the Capacitated Vehicle Routing Problem with Unsplit Demands", D. Simchi-Levi and J. Bramel

"On Refinements in Probabilistic Analysis of Geometric Problems", P. Jaillet

"Stochastic and Dynamic Vehicle Routing in the Euclidean Plane", D.J. Bertsimas and G. van Ryzin

10:15-10:45 Break / Pause-café

10:45-12:00 Frontenac

Traffic Management - Gestion de la circulation

"Smart Traffic Signals", R.C. Larson and K.V. Ballman

"A Fast Algorithm for Signal Setting at Traffic Junctions", C.A. Moyer

"A Branch-and-Bound Algorithm for the Traffic Signal Synchronization Problem with Variable Speed", P. Mireault

12:00-13:30 Lunch, then bus to Montreal / Déjeuner, ensuite départ pour Montréal

Table of Contents

Session 1
Airline Crew Scheduling – Horaires d'équipages dans l'aviation civile

Président / Chairperson:	Ranga Anbil
Date:	Thursday, June 6 Jeudi, 6 juin
Heure / Time:	09:00
Lieu / Location:	École des H.E.C. Salle / Room 3073

Optimizing the Airlines' Set-Partitioning Problems

Karla L. Hoffman* and Manfred Padberg**

* Department of Operations Research & Applied Statistics, George Mason University, Fairfax, VA 22030, U.S.A.
**Stern School of Business, New York University, New York, NY 10006, U.S.A. and
Laboratoire d'Econométrie, Ecole Polytechnique, Paris, France

---

We present a *branch-and-cut* approach to solving large set-partitioning problems and test this new algorithm on large-scale real-world crew-scheduling problems provided to us by the airline industry. This approach generates cutting planes based on the underlying structure of the polytope defined by the convex hull of the feasible integer points. The method first tightens the user-supplied formulation by automatic reformulation techniques (via row reduction, clique identification and variable fixing). Lower bounds for this problem are sequentially improved by appending additional constraints to the problem which "cut-off" fractional vertices while eliminating no zero-one points. Upper bounds are obtained by a new linear-programming based heuristic presented here. Each of these components is embedded within a tree-search algorithm called "branch and cut" where, at every node of the search tree, the problem is reformulated, linear programs are solved, polyhedral cuts are generated, and a heuristic is called.

There are four components to the *branch-and-cut* solver: a preprocessor which tightens the user-supplied formulation; a heuristic which yields "good" integer-feasible solutions quickly; a cut-generation procedure – the *engine* of this overall approach – which tightens the linear-programming relaxation, and a branching strategy that determines the direction of the search-tree. Prior research has shown that incorporating these components into one solver has successfully solved classes of problems which have been previously believed to be intractable – e.g. the symmetric traveling salesman, and general large zero-one problems. We describe how to apply this general framework to the class of problems known as set-partitioning problems.

The most widely used application of the set-partitioning model seems to be the airline crew scheduling problem, in which the rows correspond to the set of flight legs (from city A to city B, at time t) to be covered during a planning period (usually four to five days), while each column stands for a possible rotation (sequence of flight legs covered by a specified crew with the same initial and terminal point). In order to be acceptable, a rotation must satisfy a variety of regulations imposed by the federal government, airline procedures and union contracts. To set up the problem, one starts with a given set of flight legs (problems in our test set vary from 50 to 900 flight legs) and one generates by computer a set of acceptable rotations with their respective costs. Our test set consists of problems having as few as 600 rotations and as many as 300,000 rotations. The linear-programming relaxation of these problems was considered intractable until recently, and thus only heuristic methods (unrelated to linear programming technology) were applied.

The branch-and-cut optimizer described here is designed to solve *to optimality* large set-partitioning problems having thousands of zero-one variables and can alternatively be used as a heuristic procedure for obtaining "reasonable" feasible solutions and at the same time a true lower bound on the optimal solution value. The latter is, of course, a yardstick to measure the quality of the solutions found.

In this design, we begin by examining the user-supplied formulation of the problem. Within the branch-and-cut context, "preprocessing" refers to reformulation techniques which can be performed automatically at any point within the branch-and-cut algorithm to improve or simplify a given formulation.

The preprocessing component of the branch-and-cut solver takes a given $m \times n$ matrix, the "user-supplied" formulation, and automatically and iteratively reformulates the problem so that both the linear-programming relaxation of the set-partitioning problem and the zero-one problem itself are more tractable. These techniques are shown by us to be highly effective in reducing the solution times of the linear-programming subproblems and assist in obtaining the *optimal* integer solution.

The preprocessing algorithm consists of four major components which: (1) remove duplicate columns; (2) perform pairwise comparisons of rows in order to detect row inclusion. (If one row properly contains another, then all variables not in the intersection can be fixed to zero. A further extension examines pairs of rows which differ in exactly two elements. These pairwise row comparisons determine that either two columns can be "merged" since, in any feasible solution, both variables must be equal or alternatively that the two variables can both be fixed to zero.); (3) examine rows to determine if cliques exist in the intersection graph associated with the matrix A which properly contain some row k of A. (If so, we fix all variables in the clique but not in row k to zero.); and (4) we examine if after repeated applications of the above routines the intersection graph becomes disconnected, i.e. the matrix A of problem decomposes into blocks. If decomposition is successful, we then have smaller disjoint set-partitioning problems to solve.

After reformulation of the user-supplied problem, a lower bound on the problem is determined by optimizing the linear programming relaxation.

An upper bound for the problem is obtained by calling a linear-programming-based heuristic that successively sets variables to zero or one and examines the implications of such setting on other variables. It sends successively smaller problems (in terms of the number of unfixed variables) to the linear programming solver which, in turn, directs the next round of setting. Our linear-programming based heuristic exploits the empirical findings that linear-programming solution vectors of *small* set-partitioning problems are often integer even though larger problems of the same general structure are highly fractional. We therefore attempt to sequentially collapse the large problem into smaller problems. To accomplish this, we round a subset of the variables, and use the reformulation routines to detect further implications of such rounding. The linear-programming solver is called iteratively with smaller and smaller problems and the resulting solution vector directs the next stage of the rounding, feasibility checking and logical implication setting.

Given an upper and lower bound, one can use the linear-programming reduced-cost information to permanently fix variables. If the reduced-cost procedure fixes even a small proportion of the variables, such fixing can have implications for many other variables within the problem. We therefore return to the preprocessor whenever a specified percentage of the remaining variables are fixed by reduced cost fixing.

Once no more variables can be fixed and we have a good formulation of the problem, we begin the most critical phase of the solver – the constraint generation phase. These constraints are based on the polyhedral theory of integral polytopes – i.e. the theory describing the minimal complete system of linear inequalities (facet-defining inequalities) that describe the integer polytope. This phase requires not only a representation of that minimal system (or at least some reasonably large subsystem of the complete system of inequalities) but also a method for algorithmically generating such inequalities *automatically* which cut-off fractional solutions to the L.P. representation. We will present efficient implementations of the facet-identification problem for both the set-packing and the set-covering polytope (i.e. given a fractional linear programming solution, generate one or more inequalities based on the polyhedral theory that separate that solution from the integer polytope corresponding to the set-packing or set-covering problem). We begin by projecting out all non-fractional variables at their respective values in the L.P. solution and constructing the intersection graph associated with this reduced problem. We identify *cliques*, i.e. complete subgraphs, and *odd-cycles without chords* within the graph. These structures specify facets of the integer polytope for the lower-dimensional problem and can then be extended via lifting procedures so that they are valid inequalities for the fully-dimensional problem.

We note that because these polyhedral cuts are valid throughout the tree (which is not true for traditional cutting planes, such as Gomory cuts or "intersection" cuts) inequalities that are generated in *any* part of the search tree remain valid "globally", i.e. across the entire search tree, and the data structures remain unchanged when we move from one branch of the search tree to another unless the preprocessor is invoked.

The violated constraints are appended to the original problem and the L.P. solver is again called. We iterate through this loop until one of the following cases prevail: 1) The solution is integer; 2) The L.P. is infeasible; 3) No additional cuts are generated (either because of our incomplete knowledge of the polyhedral structure or because of the incompleteness of our constraint-generation procedures); 4) Although cuts are generated, the objective function is not increasing sufficiently, i.e. we detect a "tailing off" of the procedure; or 5) The objective function has improved substantially relative to the bounds on other nodes of the search tree; we "pause this node" and look at the more promising nodes.

If the first situation occurs and we are at the root node of the search tree, the solution obtained is optimal and the algorithm terminates. If this situation occurs within the search tree, we fathom the node, update the upper bound and continue the tree-search. If the second situation occurs at the root node of the tree, the overall problem is infeasible; if within the tree, we fathom the node and continue. If the third or fourth case occurs, we will expand the tree. However, before expanding the tree we call the heuristic again in an attempt to update the upper bound information. If the heuristic finds a better feasible solution, the reduced-cost fixing and logical fixing routines are called. If sufficiently many variables are fixed, the preprocessor is called again to tighten the problem formulation. The fifth case is similar to the third and fourth cases, except that the node is not expanded, but rather paused for examination at some later time.

Thus, at every node of the branch-and-cut algorithm, the problem is reformulated, linear programs are solved, polyhedral cuts are generated, and a heuristic is called.

This abstract has outlined each of the major components of the branch-and-cut solver. Details of each of the major components will be presented in an accompanying paper. This paper will also present the computational results of testing various implementations on a set of test problems provided by the airline industry and describe how this new technology has altered the way in which the airlines develop their crew schedules.

Column Generation Approaches to the Airline Crew Scheduling Problems

Martin Desrochers[1], Yvan Dumas[2], François Soumis[1] and Pierre Trudeau[3]

[1] GERAD, École des H.E.C., 5255 ave Decelles, Montreal, Canada H3T 1V6, and
Ecole Polytechnique, C.P. 6079, succursale A, H3P 3J7
[2] GERAD, École des H.E.C., 5255 ave Decelles, Montreal, Canada H3T 1V6
[3] AD OPT, 4475 St. Laurent, Suite 201, Montreal, Canada H2W 1Z8

There are several types of problems in airline crew scheduling. There are short-haul, medium-haul, and long-haul crew scheduling problems. In this paper, we present how to handle these various types of problems.

First a few definitions, a *flight leg* or leg is the portion of a flight between consecutive take-off and landing; a *duty* is the work done in one day; a *rotation* is a trip from the base to the base and composed of one or more duties. A *deadhead* leg is a leg where a crew member is assigned to travel as a passenger. It is used to transfer some crew member to a city where there is a lack of them or to return a crew member to its base.

In each problem type, we solve a set covering formulation of this problem. Each row of the set covering problem corresponds to a leg and each column to a feasible rotation. The objective is to construct a minimum cost set of rotations which covers all of the legs. As the number of columns in this formulation is very large, we need to use a column generation approach to solve it.

In all the cases, the master problem is a set covering problem with additional constraints and the subproblems are constrained shortest path problems. The main difference between the approaches used to solve the various problems is the subproblem modeling. In the case of a short-haul problem, there can be many legs per day and the number of feasible duties (one day of work) can be very big. This precludes complete enumeration of the duties. The subproblem network must be defined in terms of legs and feasible connections between legs. In the case of the medium-haul problem, the number of legs per duty is small (at most three or four) and it may be possible to enumerate all duties. In this case, the subproblem network is defined in terms of duties and night rests connecting duties. In the case of the long-haul problem, the number of legs per duty is also small (three or less) but new considerations can become important such as handling correctly deadheading, the night rests are longer, etc. We present numerical results on several types of problems.

5–11 juin/June 1991

An Alternative Solution Procedure for the Long-Haul Crew Assignment Problem

Cynthia Barnhart*, Ellis Johnson*, and Ranga Anbil**

* School of Industrial and Systems Engineering, Georgia Institute of Technology,
Atlanta, GA 30332–0205, U.S.A.
**American Airlines Decision Technologies, M.D. 3345, P.O. Box 619616, Dallas
Fort Worth Airport, TX 75261–9616, U.S.A.

Problem Description:

The inputs to the long-haul crew assignment problem include a planning horizon (e.g., one month for long-haul problems); crew base locations; and flight service, typically international, with specified departure and arrival times at fixed origin and destination points. The objective is to assign one crew to cover each flight service such that overall costs are minimized and crew availability and work rules restrictions are not violated.

Problem Formulation:

We modeled the long-haul crew assignment problem as a linear multi-commodity network flow (MCNF) problem, where each commodity represents a set of crews located at a crew base at a point in time, and the arcs of the underlying network reflect all possible legal crew assignments to flight services. The network is generated such that each network path represents a legal crew pairing, i.e., a sequence of flight services beginning and ending at a crew base without violating work rule restrictions. Thus, the MCNF objective of finding the cost-minimizing assignment of commodity flows to network paths is equivalent to determining the cost-minimizing assignment of crews to flight services.

For each flight service, two nodes and a connecting "service" arc are included in the network. The tail node represents the flight's origin location at the time of its departure and the head node represents the flight's destination location at its arrival time. The connecting service arc represents the scheduled flight service to which a crew must be assigned. To include all possible legal crew assignments in the network, additional service arcs are added, each one corresponding to a string of flight services uninterrupted by rest. These strings, often called duty periods, do not violate work rule restrictions when assigned to a single crew.

In addition to service arcs, the network contains "layover" arcs. Layover arcs, representing the possibility that a legal pairing may require crews to stay in one location for a period of time, span two nodes representing different times and a single location.

All arc costs are set equal to the time difference associated with the head and tail node of the arc. This reflects our assumption that, for long-haul crew assignment problems, the 2/7th rule will typically determine the crew assignment cost. (The 2/7th rule requires the cost of any crew assignment to be a function of the elapsed time from the departure of the crew from their base to their arrival back at the base.)

Solution Methodology:

In the arc-chain formulation of the MCNF problem, each variable represents a legal crew pairing. Thus, given our network in which each path is a legal pairing, we solve the linear crew assignment problem using column generation techniques, specifically price-directive or Dantzing-Wolfe decomposition. The first set of columns included in our linear program (LP) are determined using a heuristics we developed to generate initial feasible solutions. Solution of the initial LP provides us with dual prices used in modifying network arc costs and in generating additional LP columns. Columns for inclusion in the LP are easily generated using a shortest path procedure over the modified cost network. Further efficiencies are gained by recognizing the special time-based structure of the network allowing the shortest path algorithm to be replaced with an 0(A) network search procedure.

Computational experiences with our model, using a state-of-the-art LP solver, will be reported for a crew assignment problem encountered by a long-haul operating airline.

Contributions:

Our multi-commodity network flow modeling approach, an alternative to the classical set partitioning and heuristic approaches most often employed in practice, solves the linear long-haul crew assignment model using only the simplex and network search algorithmic tools. By generating a quality initial feasible solution, our approach succeeds in adequately restricting the size of even large-scale problems and shows the promise of producing solutions superior to those currently generated.

Session 2
Stochastic Models I – Modèles stochastiques I

Président / Chairperson:	Stein W. Wallace
Date:	Thursday, June 6 Jeudi, 6 juin
Heure / Time:	11:00
Lieu / Location:	École des H.E.C. Salle / Room 3073

Routing with Stochastic Demands: Past, Present, and Future

Moshe Dror

Decision Sciences, College of Business and Public Administration, The University
of Arizona, Tucson, AZ 85721, U.S.A.

We first review past chance-constrained and stochastic programming with recourse formulations of the VRP with stochastic demands, leading into the current research directions. The modelling aspects of this problem will be examined, and the state of the art will be described, from the perspective of searching for constructive solution procedures. We present graph-oriented interpretations of the problem, and examine solution approaches based on deterministic methodology and a Markov decision process formulation. In conclusion, research directions will be outlined for the near future and the potential of a solution process based on parallel computation machines will be discussed.

Exact Solution of Some Transportation Problems Under Uncertainty

Gilbert Laporte* and François Louveaux**

* Centre de recherche sur les transports, Université de Montréal, C.P. 6128,
succursale A, Montréal, CANADA H3C 3J7
**Facultés Universitaires Notre-Dame de la Paix, Rempart de la Vierge 8, B-5000 Namur, BELGIUM

One of the stumbling blocks of research in the field of transportation is the development of models and algorithms that explicitly deal with uncertainty in some of the factors such as the level of demand, the transportation costs or the travel time.

An important stream of research has developed, starting from early stochastic transportation models to more recent work on dynamic vehicle allocation, location on network with queues or stochastic routing problems.

Most of this research is based on methods requiring some approximation.

Laporte and Louveaux have recently developed an exact method for solving stochastic integer programs. Although general in nature, the method requires a specific tuning to each application in order to be efficient.

Such research has been already realized for routing with stochastic travel time or stochastic demand under limited vehicle capacity. It has also been realized for location problem under uncertain demands.

The object of the presentation will be to review some of the available implementations and to report on the results of the numerical experiments to emphasize the strength and limitations of the method.

It will also be to present some of the more recent research developments related to the transportation problems.

Stochastic Programming Applications in Transportation: Dynamic Network Models of Common Carrier Operations

Warren B. Powell*, Raymond K.L. Cheung* and Linos F. Frantzeskakis**

* Department of Civil Engineering & Operations Research, Princeton University, Princeton, NJ 08544, U.S.A.
**Department of Operations Research, Room 3K-320, AT&T Bell Laboratories, Crawfords Center Road, Holmdel, NJ 07733, U.S.A.

Dynamic network models represent a powerful framework for modeling common carrier operations. These can be used to study dynamic fleet management problems (empty repositioning, load acceptance and load solicitation), as well as the service network design problem, fleet planning, and costing studies. In a dynamic (rolling horizon) setting, an important technical problem is handling forecasting uncertainties. We review models that represent these problems as networks with random arc capacities, and survey different methodological approaches for approximating recourse functions in stochastic programming, based on the concept of restricted recourse strategies. These are used to develop analytical approximations of recourse functions that can be used to optimize stochastic programs.

Session 3
Focus 1

Président / Chairperson:	Warren B. Powell
Date:	Thursday, June 6 Jeudi, 6 juin
Heure / Time:	14:00
Lieu / Location:	École des H.E.C. Salle / Room 3073

New Technology for Transportation

David K. Willis

The ATA Foundation Inc., 2200 Mill Road, Alexandria, VI 22314–4677, U.S.A.

The presentation will examine the potential utility of so-called "intelligent vehicle-highway systems" (IVHS) technologies to reduce traffic congestion, enhance commercial vehicle productivity, and increase highway safety. Technologies to be addressed include: advanced traffic management systems; "real-time" traffic information services; automatic vehicle location (AVL) and two-way communication technologies; automatic vehicle identification (AVI); routing algorithms; and automatic vehicle control technologies.

Session 4
Equilibrium Models – Modèles d'équilibre

Président / Chairperson:	Michael Florian
Date:	Thursday, June 6 Jeudi, 6 juin
Heure / Time:	15:15
Lieu / Location:	École des H.E.C. Salle / Room 3072

An Interactive Spatial Price Equilibrium Model for Designing Long-Run Strategies for Competitive Firms

Antonio Edmundo de Rezende

COOPE/UFRJ, Centro de Tecnologia, B1.H,s/117, Caixa Postal 68512, 21945 Rio de Janeiro, RJ, Brazil

Firms make the long-run decisions on both location and size of production and transportation facilities vis-a-vis their expected market demand. The model of spatial price equilibrium (MSPE) simulates short-run market decisions and indicates the market prices and the product flows for the whole industry, given production and transportation capacity. The investment alternatives are generated by every firm as it evaluates its market potentials, by appreciating future transportation cost and competition. The firm's investment plan is justified on the expected series of discounted revenues minus cost.

The MPSE transforms the system of production, transportation and consumption into a (single commodity) network flow problem. Each production site is represented by its supply function, each transportation link is represented by its congestion curve and each consumption point by its inverse demand function. A series of parallel arcs represent a curve. These curves are assembled in a network that represents one market period. Every planned investment is a curve added to the network or dropped from it when some existing facility is being closed. The market (partial) equilibrium is then obtained by minimizing the MPSE flow cost which is equivalent to maximizing the social surplus.

The simulation of logistics strategies is made for a number of periods and requires that all the firms in the industry be maximizing profits. Each firm may increase its profit by investing in productive capacity along with alternatives of distribution, as can be the case, in order to reduce fob prices. The manager can combine plant capacity and freight terminals to configurate distribution strategies and can then simulate the results with the model. Expected demand expansions as well as competitors' expected investments shall also be incorporated in the market scenario of the firm. The analyst will generate as many (market) networks as the number of periods of study. Series of financial results are then obtained from the simulation and are used to evaluate the firms' market performance.

The accuracy of the equilibrium depends on the number of parallel arcs of each curve in the network. The out-of-kilter algorithm solves the MPSE network efficiently. Reasonable large problems (two thousand arcs) were run in the simulation of market experiments with six or eight participants (firms) controling over 25 plants and freight terminals, 50 sites of demand and over 1200 transportation links. The MPSE model and the dynamics of the experiment are described in the paper. Noncompetitive equilibrium will also be discussed in the context of the experiment.

The MSPE yields partial equilibrium for a single commodity market. However, as the model has the ability to transform nonlinear supply, demand and congestion curves into a linear network flow problem, it is feasible to solve quite large market problems within reasonable computational time in a PC-XT. Nonlinear algorithms to solve SPE are applied to small markets only. The MSPE linear approximation to supply, demand and congestion curves, which depends on the number of parallel arcs (steps) in the curve, can therefore be made as good as desired without computational problems.

The model has been applied to simulate market experiments with students of logistics. They are supposed to make long-run investment decisions at every period, based on their expectation on the growth of demand (whose random behavior is known) and on their appreciation about the competitors' most likely decisions. Then, they locate and dimension facilities in a competitive environment and anticipate the financial returns for the investment. MSPE gives them market short-run prices and flow at each period. So, later on the experiment they will be able to verify their forecasts against the market results. The experiment stresses the choice of channels of distribution (plant location and dimension, railroad capacity, freight terminal location) to give the firm better market access and share. In sum, the decision maker (student) faces how to integrate demand growth, competition, production and transportation costs in effective formulation of logistic strategies. That is the MSPE relevant contribution.

Supply Demand Equilibrium on Transportation Networks with Composite Modes

Enrique Fernandez*, Joaquin de Cea* and Michael Florian**

* Departamento de Ingeniería de Transporte, Pontificia Universidad Católica de Chile, Casilla 6177, Santiago, Chile
**Centre de recherche sur les transports, Université de Montréal, C.P. 6128, succursale A, Montreal, Canada H3C 3J7

Important advances have been achieved during the last ten years in the formulation, understanding and analysis of the resulting models and development of algorithms for computing multimodal equilibria over transportation networks. Some of the most relevant contributions in this field are mentioned in the list of references at the end of this abstract.

Nevertheless, the models most often formulated and studied, employ the concept of a generalized abstract mode (3) or explicitly consider only pure modes (5). The possibility of "composite modes"[1] has never been explicitly formulated in a general supply-demand multimodal network equilibrium context, even though many demand models have been specified and calibrated with their explicit consideration. A partial approach to the modelling of such modes is presented in (6) within the context of the management of parking lot facilities.

Composite modes are of obvious importance in any practical study, given that most current policies, especially those related to urban transportation, tend to increase their share in the transport market, such as "park and ride" or integrated transit systems using different combinations of public transport modes.

The existence of composite modes creates special modelling challenges requiring a one-to-one correspondence between options considered in the demand models used and the corresponding structure of the network formulation. If a composite mode is treated as a different option in the demand model, then a special network must correspond to that "composite mode" in order to properly model route choice as well as the interactions with the other modes present in the multimodal network considered.

A proper solution to modeling "composite modes" must provide answers to the following conceptual questions: What traveler decisions must be modeled by the demand models and which by the network models? Is a composite mode a different mode in itself, with a different attractiveness and subjective valuation by the passengers or just a routing combination of pure modes on the network? How should the costs (disutility) considered by the demand models be defined in order to assure compatibility with parallel measures used for network models? What is the effect of considering different demand models? (i.e. multinomial logit vs. hierarchical logit). What changes are necessary in the known analytical formulation of network equilibria in order to explicitly include composite modes? How will these changes affect the mathematical properties of the problem? The paper addresses these questions, provides a general framework for the consideration of composite modes and identifies topics both theoretical and empirical which require additional research.

---

[1] We call composite mode to a combination of pure modes. A trip on a composite mode has then different sections, each of them incurred over a different modal network and linked by transfers, i.e. car-metro, bus-metro, etc.

References

1. Aashtiani, H.Z. and T.L. Magnanti, "Equilibria on a Congested Transportation Network", *SIAM J. Algebraic Discrete Meth.* 2, 213–216 (1981).

2. Abdulaal, M. and M.J. Leblanc, "Methods for Combining Modal Split and Equilibrium Assignment Models", *Trans. Sc.* 13, 292–314 (1979).

3. Dafermos, S.C., "The General Multimodal Network Equilibrium Problem with Elastic Demand", *Networks* 12, 57–72 (1982).

4. Florian, M., "A Traffic Equilibrium Model of Travel by Car and Public Transit Modes", *Trans. Sc.* 8, 166–179 (1977).

5. Florian, M. and H. Spiess, "On Binary Mode Choice/Assignment Models", *Trans. Sc.* 17(1), 32–47 (1983).

6. Florian, M. and M. Los, "Determining Intermediate Origin-Destination Matrices for the Analysis of Composite Mode Trips", *Transp. Res. V.* 13B(2), 91–103 (1979).

Modeling and Solving the Dynamic Traffic Assignment Problem with Windows of Time

Omar Drissi-Kaïtouni, Ali El-Merzouqi and Michael Florian

Centre de recherche sur les transports, Université de Montréal, C.P. 6128,
succursale A, Montreal, Canada H3C 3J7

This paper is concerned with the models aimed at simulating the flows of urban transportation networks during the peak periods. These models, that are referred to as Traffic Assignment Models (TAM) in the literature, may be classified into two main classes that are: the Static and the Dynamic Traffic Assignment Models (respectively STAM and DTAM).

In the first class of models a static demand for a time period is assigned on the network in order to minimize the individual transportation costs (user optimal assignment). In spite of the fact that the demand is not constant during the peak period in an urban context, the STAM's are extensively used in practice.

Dynamic Traffic Assignment Models are more suitable than the static ones, to analyze the flow variations on the links during the peak periods, since the demand is allowed to vary in time. Unfortunately, most of the models in this class proposed in the literature are too restrictive. Some are heuristic models and do not express a descriptive behaviour of the users, and other consider only parallel path networks where congestion occurs at a single bottleneck link.

However, since 1978, many optimization (and optimal control) models have been proposed in the literature for the DTAM. In 1978, Merchant and Nemhausser present a first minimization model for the single origin-destination DTAP. Ho (1980) proposes a stepwise linear version of this model which has been more deeply investigated by Carey (1986 and 1987); this author presents a convex version of the model and some techniques to generalize it to many origins and many destinations. Finally, Friesz et al. (1990) proposes a continuous time version of the model (optimal control problem). In these models, the dynamics of link flows are modeled by using link exit functions, that are restricted to be concave in order to obtain convex programs. Other restrictive conditions are imposed on the link cost functions in order to insure that the flows reach the destination. Finally, these models are expressed in terms of the link flows for one destination, and do not lend themselves to natural extensions for the multicommodity case (many destinations).

A few years ago, Zawack and Thompson (1987) adapted for a DTAM the network model proposed for the Dynamic Maximum Flow Problem by Ford and Fulkerson (1962) and Fulkerson (1975). In this model, each link $a \in A$ has a capacity and a fixed traversal time. For any period of time, at each node, vehicles can either be transshipped on exit links or wait until the next period. The resulting model is a transshipment problem over an expanded (space-time) network, where only one destination is allowed.

Recently, Drissi and Hameda (1989) formulated this model in terms of path flows (in the space-time expanded network) as a multicommodity network equilibrium problem, where the link capacity constraints are incorporated in the objective function as penalty terms. Further, they modeled the expanded space-time network in a way to include the queues on the links rather than on the nodes.

In this paper, we consider a variant of this model which is formulated over a space-time network by considering as well desired "time windows" of arrival at destinations. We show that the departure times of the users may be easily handled by this model by setting penalties at the origin and destination nodes, for early departures and late arrivals.

We then show that this problem may be solved efficiently by the linear approximation (LA) algorithm and its variants (e.g., Partan) as well as by other algorithms such as the Truncated Quadratic Programming algorithms that may use the LA algorithm (or Partan) to solve their own subproblems.

Finally, we show the relation between this DTAM and the standard STAM.

Numerical results obtained with the transportation networks of the cities of Hull and Winnipeg (Canada) are presented.

A Conjugate Gradient Algorithm for the Dynamic User Optimal Traffic Assignment Problem

Byung-Wook Wie

School of Travel Industry Management, University of Hawaii, George Hall, 2560
Campus Road, Honolulu, HI 96822, U.S.A.

The conjugate gradient method is applied to solve a discrete-time version of the dynamic user optimal traffic assignment problem that was previously analyzed in Friesz, Luque, Tobin and Wie (1989). The solution algorithm is illustrated by a numerical example. Numerical results show that the algorithm converges to an optimal solution satisfying the instantaneous dynamic user equilibrium conditions. These conditions are referred to as a dynamic generalization of Wardrop's first principle such that if, in each time period, for each origin-destination pair, the instantaneous unit cost, the corresponding time-varying traffic flow pattern is said to be user optimized.

References

Friesz, T.L., F.J. Luque, R.L. Tobin and B.W. Wie (1989) "Dynamic Network Traffic Assignment Considered as a Continuous Time Optimal Control Problem", *Operations Research* 37(6), 893–901.

Session 5
Location and Networks – Modèles de localisation

Président / Chairperson:	Margaret L. Brandeau
Date:	Thursday, June 6 Jeudi, 6 juin
Heure / Time:	15:15
Lieu / Location:	École des H.E.C. Salle / Room 3074

Dynamic Facility Location Modelling Using Forecast Horizon Concepts: Issues, Algorithms and Results

Mark S. Daskin* and Wallace J. Hopp**

* Department of Civil Engineering, Northwestern University, Evanston, IL 60208, U.S.A.
**Department of Industrial Engineering, Northwestern University, Evanston, IL 60208, U.S.A.

Facility location modelling has been the subject of intensive research for many years. Most location models are static; that is, they consider only a single time period. In fact, location decisions are made and implemented over a period of time and the facilities that are sited are designed to serve demands over an even longer time frame. During the time that facilities are to be operational, demands, travel times, costs and other input factors are likely to change significantly. Consequently, a dynamic approach to facility location modelling and decision making is called for.

Attempts to capture the dynamic nature of the facility location problem fall into two broad categories: (1) temporal extensions of static modelling techniques and (2) scenario planning approaches. Temporal extensions of static techniques typically involve appending a time dimension to static models. The location set covering model and the maximum covering location model (Chrissis, Davis and Miller, 1982; Gunawardane, 1982; and Schilling, 1980) as well as the P-median model (Wesolowsky and Truscott, 1975) and the fixed charge facility location model (Van Roy and Erlenkotter, 1982) have all been adapted to dynamic contexts in this manner. In all such approaches, the time horizon over which the models extend is an *exogenous* input. The key problem that this creates is that if the planning horizon is too short, myopic initial decisions will be made that lead to long-term suboptimal locations. On the other hand, if the planning horizon is too long, excessive data collection is required and solution times are likely to grow, thereby inhibiting the ability of planners and decision makers to use the models in "what-if" planning modes. In addition, the solution to such models is an "optimal" set of decisions for *all* periods during the planning horizon. This, in fact, is more than is needed in most cases. What is actually needed is only an optimal *first-period* decision since plans are revised frequently as new information is obtained.

Scenario planning approaches are designed to account for the uncertain nature of future conditions. In such an approach, the model is run against a number of future conditions. A solution that is good with respect to all future conditions is sought. Schilling (1982), for example, suggested implementing those facility sites that are common to all scenarios. Medina (1989) and Daskin, Hopp and Medina (1990) showed that for some problems this approach will result in the adoption of the *worst* possible first period decision.

Medina (1989) and Daskin, Hopp and Medina (1990) introduced the notion of forecast horizons in location modelling. Forecast horizons capture two key features that are missing in current dynamic facility location models. First, in using a forecast horizon approach, the analyst seeks only an optimal first period decision. Second, the length of the planning horizon is determined *endogenously* to be the shortest planning horizon such that the initial decision is invariant with respect to conditions beyond the end of the planning horizon. Daskin, Hopp and Medina (1990) showed that an optimal forecast horizon may not exist in location problems. Consequently, they introduced the notion of an **ϵ–optimal forecast horizon** and an **ϵ–optimal first period decision.** Using DYNALOC (Van Roy and Erlenkotter, 1982), they showed empirically that ϵ–optimal forecast horizons and first period decisions are likely to exist for small values of ϵ.

The purpose of this paper is to report on more recent theoretical and methodological work related to ϵ–optimal forecast horizons and ϵ–optimal first period decisions in location modelling. Specifically, we will show how a variety of static location models can be extended to the dynamic location regime using forecast horizon concepts. In addition, we will discuss preliminary algorithms for the fixed charge dynamic facility location problem that (1) seek only an ϵ–optimal first period decision (while bounding the costs incurred in subsequent periods) and (2) determine the length of an ϵ–optimal forecast horizon endogenously. Using a number of small data sets we will explore the sensitivity of the solution time and the length of the planning horizon to the value of ϵ, the number of nodes in the network, as well as other key inputs. Directions for future study will be outlined.

References

Chrissis, J.W., R.P. Davis, and D.M. Miller (1982) "The Dynamic Set Covering Problem", *Applied Mathematical Programming* 6, 2–6.

Daskin, M.S., W.J. Hopp and B. Medina (1990) "Forecast Horizons and Dynamic Facility Location Planning". Submitted to *Annals of Operations Research* (also presented at the 1990 ISOLDE V Conference, Fullerton, CA, U.S.A.)

Gunawardane, G. (1982) "Dynamic Versions of Set Covering Type Public Facility Location Problems", *European Journal of Operations Research* 10, 190–195.

Medina, B.B.V. (1989) "Dynamic Facility Location Modelling", M.S. Thesis, Department of Civil Engineering, Northwestern University, Evanston, IL 60208, U.S.A.

Schilling, D.A. (1980) "Strategic Facility Planning: The Analysis of Options", *Decision Sciences* 13, 1–14.

Van Roy, T. and D. Erlenkotter (1982) "A Dual Based Procedure for Dynamic Facility Location", *Management Science* 22, 1363–1373.

Wesolowsky, G.O. and W.G. Truscott (1975) "The Multiperiod Location-Allocation Problem with Relocation of Facilities", *Management Science* 22, 57–65.

On Locating Path- or Tree-Shaped Facilities on Networks

S.L. Hakimi[1], E.F. Schmeichel[2] and Martine Labbé[3]

[1] Department of Electrical Engineering and Computer Science, University of California, Davis, CA 95616, U.S.A.
[2] Department of Mathematics and Computer Science, San Jose State University, San Jose, CA 95192, U.S.A.
[3] Econometric Institute, Erasmus University, Box 1738, 3000 DR, Rotterdam, THE NETHERLANDS

The study of "optimality" locating on a network a single facility of a given total length in the form of a path or a tree was initiated by several authors. We extend these results to the problem of locating $p(\geq 1)$ such facilities. We attempt to provide a complete complexity analysis of "center", "median", "max eccentricity", and "max distance sum" location type problems for $p = 1$ or $p > 1$, for general networks and for tree networks, whether a facility contains partial arcs or not, and whether a facility is path-shaped or tree-shaped. These cases lead to 64 problems, of which 31 are NP-hard, 29 are solvable in polynomial time and 4, all max distance sum problems, have complexities which remain to be determined. We conclude with a result that may be viewed as a generalization of the p-Median theorem.

A Tabu Search Heuristic for Location-Allocation Problems with Balancing Requirements

Michel Gendreau*, Teodor Gabriel Crainic, Patrick Soriano* and Michel Toulouse***

* Centre de recherche sur les transports, Université de Montréal, C.P. 6128,
succursale A, Montréal, CANADA H3C 3J7

**Département des Sciences administratives, Université du Québec à Montréal, and Centre de recherche sur les transports, Université de Montréal, C.P. 6128, succursale A, Montréal, CANADA H3C 3J7

The strategic and tactical planning of the land transportation of empty containers involves the selection of a number of depots to which empty containers are sent by import customers and from which they are dispatched to export customers, the allocation of customers to these depots for each type of container and direction of movement, and the determination of interdepot balancing flows which account for the regional imbalances between imports and exports. This problem can be formally stated as a multimode multicommodity location-allocation problem which is a mixed integer program with integer location variables and continuous flow variables. An interesting feature of this formulation is that once the integer location variables have been fixed, the "optimal" values of the continuous variables can be obtained by solving a large capacitated multicommodity flow problem, which is equivalent to a series of one-commodity minimum cost flow problems (one per container type). It is this feature which makes it possible to adapt Tabu Search, a technique usually used to solve pure integer problems, to the problem at hand.

Generally speaking, Tabu Search, which was first introduced by F. Glover, is a local search method which at each step moves from a point in the solution space to its best (with respect to the objective) neighbour. It differs from other local search heuristics in that it does not terminate when a local optimum is encountered; instead, the search continues by executing moves in which the objective function is allowed to deteriorate. Cycling is prevented by a "tabu mechanism" which inhibits the reversal of the most recent moves.

In this presentation, we describe a Tabu Search heuristic for the location-allocation problem with balancing requirements. The solution space explored by the search is defined by the integer location variables. Although this makes the search process much more efficient than a search in the joint space of integer and continuous variables, it also makes it much more difficult to correctly evaluate the objective function in the neighborhood of a given solution. To circumvent this difficulty, a surrogate objective, which can be evaluated without solving minimum cost flow problems, is introduced to direct the search. Recent refinements in Tabu Search such as "search intensification" and "search diversification" have also been implemented in the proposed heuristic. They will be explained in detail and discussed.

Computational results on several problems will be reported and compared to those obtained when using alternate solution methods.

Session 6
Freight Transportation Models – Modèles pour le transport des marchandises

Président / Chairperson:	Uwe Pape
Date:	Thursday, June 6 Jeudi, 6 juin
Heure / Time:	16:45
Lieu / Location:	École des H.E.C. Salle / Room 3072

A Model, Algorithms and Strategy for Train Pathing and Planning

Malachy Carey

Department of Statistics, University of Oxford, 1 South Parks Road, Oxford, OX1 ETG, ENGLAND

This paper is concerned with developing and solving optimization models of the train pathing (and planning) problems for relatively general rail networks. The train pathing problem is the problem of assigning trains to lines and platforms so as to minimize travel times or costs while satisfying the various constraints on departure/arrival times, speeds, headways, wait times, etc.

Traditionally, train pathing was done by manual graphical methods, by constructing time-distance diagrams by hand. In some cases computer database, spreadsheet and graph plotting aids have been introduced to facilitate and speed up input and output and hence the whole process. However, these aides do not resolve pathing conflict, do not make trade off's or choices and do not actively search for solutions. This is still left to the human train pather/planner.

More recently, various authors have developed optimization models for the train pathing problem. Sauder and Westerman (1983) describe a train pathing system based on a relatively simple optimization model to solve the train "meet pass" problem for a single track line. Jovanovic and Harker (1988), Kraay, Harker and Chen (1988) have developed a much more comprehensive set of models. Again their model and algorithms generally assume a single track (with sidings), which is the norm for N. American freight railroads. They develop special fast algorithms to take advantage of this problem structure.

Here, however, we are concerned with more general more complex rail networks which are usual in UK and European passenger railways. For exemple, in these railways,

(a) there are usually two (or more) lines in each direction between stations and often multiple platforms at stations,
(b) trains often have a choice of lines between stations and platforms at stations,
(c) headways are often very short (e.g., less than 3 or 4 minutes),
(d) in some cases trains can choose to bypass a station altogether,
(e) there may be complex junctions or intersections (with headway requirements) at station entrances/exits and elsewhere,

and so on.

We develop a model which includes all of the above features. This yields a large scale 0–1 integer programming problem, and is not solvable using general purpose packages or algorithms. In view of this we develop special algorithms which take advantage of the problem structure. These algorithms are intended to be,

(a) flexible enough to handle a wide variety of network layouts which occur in practice for passenger railways, especially in the UK and Europe, and
(b) fast enough to solve realistic sized problems quickly in practice.

We achieve this partly by emulating some problem decomposition strategies used by expert train planners in practice. A basic step (repeated many times) in our algorithms is to "path in" a single train by (a) letting its sequence order relative to other trains vary while (b) holding the path order of other trains fixed relative to each other and (c) allowing the times of some or all trains to vary.

The model and algorithms also allow for different costs/values of times for different trains, different lines, etc., and allow the introduction of reliability costs, based on weighting the time intervals between trains. The model and algorithm can also be extended in other ways.

In computer runs so far (on a SUN workstation) we have found that initial versions of the algorithms solve problems fast enough (in a few minutes) to be useful in practice. Solution times increase only lincarly with the number of trains, if the train density (service frequency) is not increased. (A price of increased generality and flexibility may be that the algorithms may not be as fast as those of Jovanovic, Harker and Kraay when solving the special class of problems which they consider).

References

Jovanovic, D. and P.T. Harker (1988), "Railroad schedule validation and creation: the Scan I system", forthcoming in Transportation Science.

Kraay, D., P.T. Harker and B. Chen (1988), "Optimal pacing of trains in freight railroads: model formulation and solution", forthcoming in Operations Research.

Sauder, R.L. and W.M. Westerman (1983), "Computer-aided train dispatching: decision support through optimization", Interfaces 13(6), 24–37.

The Intermodal Trailer Assignment Problem: Models, Algorithms, and Heuristics

Thomas A. Feo* and Jose Luis Gonzalez-Velarde**

* Operations Research Program, Mechanical Engineering Department, University of Texas at Austin, Austin, Texas 78712–1063, U.S.A.
**Centro de Sistemas de Manufactura, Instituto Tecnologico de Monterrey, Monterrey, N.L. 64849, MEXICO

The problem of optimally assigning highway trailers to railcar hitches in intermodal transportation is studied. The importance of the problem is evidenced by the fact that a small improvement in operating efficiency, such as a one percent increase in hitch utilization, translates into several hundred thousand dollars in cost savings per year for an intermodal rail yard.

Two different integer-linear programming formulations are constructed. The first is based on the explicit dimensions of the equipement and yields a large optimization program. It is solved via a branch-and-bound approach employing Lagrangean relaxation and subgradient optimization to find lower bounds, and a simple 2–exchange heuristic to obtain upper bounds.

The first formulation is onerous to solve, and this burden motivates the need for a different approach. The problem is cast as a set cover in the second formulation. Given an enumeration of the different types of trailers that can be loaded onto the different types of railcars, the formulation finds a minimum cost set of railcars that cover (i.e. accommodate) all trailers. The second formulation is very small in comparison to the first, and possesses a tight linear programming relaxation in practice. This allows it to be solved by a general purpose branch-and-bound code.

The second formulation also provided the basis for the development of a Greedy Randomized Adaptive Search Procedure (GRASP). This heuristic is observed to be extremely fast. Empirically, if finds the optimal solution to all of the problem instances furnished over a two year period by Consolidated Rail Corporation. These solutions suggest that within little effort, up to a ten percent improvement in hitch utilization may be attainable in practice.

Session 7
Routing and Scheduling I – Routes et horaires I

Président / Chairperson:	Pierre Dejax
Date:	Thursday, June 6 Jeudi, 6 juin
Heure / Time:	16:45
Lieu / Location:	École des H.E.C. Salle / Room 3074

A Two-Level Model for Helicopter Routing and Scheduling Solved by Column Generation

Jacques Desrosiers[1], Yvan Dumas[1], Marius M. Solomon[2] and Alain Chalifour[3]

[1] GERAD, École des Hautes Études Commerciales, 5255 avenue Decelles, Montreal, Canada H3T 1V6
[2] Northeastern University, 314 Hayden Hall, Boston, MA 02115, U.S.A.
[3] Université du Québec à Trois-Rivières, Trois-Rivières, Québec, Canada G9A 5H7

In this paper we address the problem of routing and scheduling a fleet of helicopters which, departing from the same aerial base, undertake circuits spanning several days and consisting of visiting a number of locations. This problem arises in the context of large control programs involving the treatment with larvicides of large numbers of adult black flies breeding sites dispersed over vast areas.

In the problem of larvicide transport by helicopter, we seek to minimize the transportation costs incurred by visiting all the injection points subject to constraints on the number of helicopters, fuel tank capacity, larvicide tank capacity, flying time per day, fuel and replenishment depot locations, and overnight city locations. This problem is a difficult variant of the vehicle routing problem involving multi-day, pickup and delivery, and multi-commodity characteristics.

Due to its computational difficulty, this problem is separated in two phases. The first phase involves the grouping of nearby injection points along a river to form a set of tasks. This geographical mini-clustering phase takes optimization decisions considering only local constraints such as fuel and larvicide tank capacities, and fuel and replenishment depot locations. The second phase of the solution strategy deals with global optimization decisions over several day circuits.

The integer programming formulation at the second phase essentially involves two types of decision variables and two levels of constraints. At the first level, mathematical relations describe constraints on single day trips. These trips start and end at feasible pairs of overnight cities and each binary variable covers a subset of the tasks. At the second level, constraints are given on circuits of at most three days; they also link together trips over tasks and circuits over cities. Circuits start and end at the central air base and only cover overnight cities. They are associated to nonnegative integer variables. The first level of constraints from a set covering type model involving the binary trip variables. The second level is a flow conservation type model.

A column generation approach can be used to solve this two-level model. Trip columns are generated using a constrained shortest path algorithm on several sub-networks, each of them having an overnight city as sink and source, plus depot location nodes. Tasks are represented by two nodes linked by arcs that can be traversed in both directions. Circuit columns may be generated using a classical shortest path algorithm on an acyclic network having the air base as source and sink, and the duplicated set of overnight cities as the first and second nights. They can also simply be enumerated at the start.

The GENCOL software for vehicle routing and crew scheduling has been adapted to solve this problem. Detailed model and implementation strategies as well as numerical results on an actual application in Western Africa will be presented.

A Matching Based System for Solving a Special Class of Routing Problems

Ulrich Derigs and A. Metz

Lehrstuhl für Betriebswirtschaftslehre VII, Universität Bayreuth, Postfach 3008, 8580 Bayreuth, GERMANY

The classical vehicle routing problem (VRP) involves the routing of a fleet of vehicles of known capacity from the central depot to a set of customers with known demands under the objective to minimize the total cost or distance travelled over all routes. Here all routes must originate and terminate at the depot and the total demand on any route must not exceed the capacity of the vehicle assigned to service that route. The vehicle routing problem with time window constraints (VRPTW) is a variation of VRP where the service for each customer has to be scheduled within one or more time windows which are associated with every customer.

In this paper we deal with the following special variation of VRPTW: Given a nonhomogenous fleet of vehicles and a fixed set of customers, during one time period, i.e. a day or a week, these customers have to be delivered in the "first half" of the period with a certain amount of goods. Thereby delivery may start at time t_{start} say at the depot and for every customer there is a so-called cut-off time for the latest possible delivery. In addition to travel time there is a certain delivery-time associated with every customer, i.e. an estimated time span the vehicle has to spend at the customer location when delivering.

In the "second half" of the time period the vehicles have to pick up certain amounts of goods and to ship them to the depot. Again there is a cut-off time for the earliest possible pickup and a certain time span consumed for every pickup. The entire pickup has to be completed at a certain cut-off time, i.e. the latest possible time t_{end} say, when all vehicles have to be back at the depot.

"Driving" from the depot to the customer and driving between customers is associated with a transportation cost and a transportation time which may vary with the type of vehicle. The problem is then to design a delivery/pickup plan which meets the time and capacity requirements and which has minimal cost. Problems of this type arise for instance when data-carriers or printed output has to be exchanged between local banks and a DP-service center, when film material from local supermarkets is developed and processed in a central lab and in the express mail business. In the latter case from a certain number of airports collected express mail is flown into a central hub where all the incoming mail is sorted and then flown out to their respective destination airports.

In the following we will show that this problem can be formulated as a (high dimensional) set-partitioning problem with two additional nontrivial sets of side-constraints, a problem which is by no means efficiently solved in general. Yet, assuming that the number of customers that can be served by a single vehicle on a delivery or pickup-pass is at most two, the problem reduces to a matching problem with side-constraints. Although the problem is still NP-complete it become practicable in the sense that by relaxation and applying effective optimization techniques from nonsmooth optimization and efficient matching software good approximate solutions are constructed in acceptable time. This assumption, though relatively strong in general, is feasible in the case of the "overnight-mail" problem due to the time consumed by starting and landing and the delivery and pickup time windows at the customers.

Session 8
Transit Network Models – Modèles de réseaux de transport en commun

Président / Chairperson:	Sang Nguyen
Date:	Friday, June 7 Vendredi, 7 juin
Heure / Time:	09:00
Lieu / Location:	École des H.E.C. Salle / Room 3073

A Variational Inequality Model for Transit Equilibrium Assignment

Jia Hao Wu*, Michael Florian* and Patrice Marcotte**

* Département d'informatique et de recherche opérationnelle and Centre de recherche sur les transports, Université de Montréal, C.P. 6128, succursale A, Montréal, CANADA H3C 3J7

**Collège militaire royal de St-Jean, Richelain, Quebec, CANADA J0J 1R0 and Centre de recherche sur les transports, Université de Montréal, C.P. 6128, succursale A, Montréal, CANADA H3C 3J7

We analyze in this paper algorithms for solving the transit equilibrium assignment problem (TEAP) on transit networks that are subject to congestion. The behavior of the transit users is modelled by using the concept of strategies or hyperpaths. The formulation is applied to a conceptual network that results from the transformation of a physical transit network composed of walk, wait, transfer/alight and in-vehicle arcs, with corresponding cost functions. The waiting cost (time) is influenced by both the interarrival times of transit vehicles and the users' queues at bus stops.

The TEAP is formulated as a variational inequality and solved iteratively by a linearized Jacobi and a projection method.

We show that the strong monotonicity of the cost function in the space of link flows (and not in the space of hyperpaths) is sufficient to obtain convergence and uniqueness results for our model.

We prove, under some usual sufficient conditions, that the linearized Jacobi method is globally convergent. The implementation of the algorithm and computational experiments are also presented.

Finally, we compare this model to another formulation of the TEAP due to Fernandez and de Cea (1989) and highlight the advantages of our approach.

A Method for Optimizing the Frequencies in a Transit Network: A Special Case of Nonlinear Bi-Level Programming

Isabelle Constantin and Michael Florian

Département d'informatique et de recherche opérationnelle and Centre de recherche sur les transports,
Université de Montréal, C.P. 6128, succursale A, Montréal, CANADA H3C 3J7

The optimization of frequencies is an important step in the design of a transit network, yet it is not well integrated in the traditional models used in the transit planning process. Until now, it has received scant attention in the literature and has often been studied in conjunction with the design of transit line itineraries. Furthermore, none of the models proposed so far (see e.g. Hasselström [1981] and Schéele [1977]) take explicitly into account the travelers' behavior regarding route choice.

We define the optimal frequencies as the ones that minimize the total expected travel time on the network, while satisfying a fleet size constraint as well as lower bound constraints on the line frequencies. We formulate this problem as a mixed integer nonlinear program with an ill behaved objective function, which is neither convex nor concave, and binary variables. We show that it can be re-written as a MIN-MIN nonlinear bi-level program.

The theory of Danskin [1967], developed for MAX-MIN problems, provides a very good theoretical framework for our problem. In this context, the minimization at the upper level is performed on the frequencies. The lower level problem turns out to be a transit assignment problem with fixed frequencies. This problem was studied by Spiess [1984], and also reported in Spiess and Florian [1989]. This model is based on the concept of "optimal strategies", and can be solved very efficiently by a polynomial time label-setting algorithm, which is similar to a shortest path label-setting algorithm.

The solution method we propose for the optimization of frequencies is of the gradient projection type (see Luenberger [1984]). Since the objective function is not differentiable continuously (discontinuity points correspond to frequencies for which the optimal strategy is not unique), the gradient does not exist everywhere. This is the reason why we project a *sub-gradient* instead of the gradient. A sub-gradient (corresponding to a particular strategy) for the upper level problem can easily be computed while solving the transit assignment problem. It is justified theoretically by using a sensitivity analysis of the lower level problem. The particular structure of the problem permits the determination of the projected direction analytically (instead of computing it via several matrix products). The optimal step is determined by using an Armijo/Goldstein criterion (see Luenberger [1984]). We prove that this algorithm converges towards a local minimum.

The main computational effort of our method is the solution of a sequence of transit assignment problems with fixed frequencies. Since this problem can be solved in polynomial time, our method can be applied to large-scale networks.

Finally, we present numerical results for two real networks.

References

Danskin, J.M., *"The Theory of Max-Min and its Application to Weapons Allocation Problems"*, Springer-Verlag, New York (1967).

Hasselström, D., "Public Transportation Planning: A Mathematical Programming Approach", Ph.D. thesis, University of Gothenburg, Sweden (1981).

Luenberger, D.G., *"Linear and Non-linear Programming"*, Addison-Wesley, Reading (1984).

Schéele, S., "A Mathematical Programming Algorithm for Optimal Bus Frequencies", Ph.D. thesis, University of Linköping, Sweden (1977).

Spiess, H., "Contributions à la théorie et aux outils de planification des réseaux de transport urbain", Ph.D. thesis, Département d'informatique et de recherche opérationnelle, Université de Montréal (1984).

Spiess, H. and M. Florian, "Optimal Strategies: A New Assignment Model for Transit Networks, *Transportation Research B*, 23B(2), 83–102 (1989).

Route Generation and Improvement Algorithms for the Transit Network Design Problem

M. Hadi Baaj* and Hani S. Mahmassani**

* Department of Civil Engineering, Arizona State University, Tempe, AZ 85202, U.S.A.
**Department of Civil Engineering, University of Texas at Austin, Austin, TZ 78712, U.S.A.

The focus of the research is the transit network design problem (TNDP) that arises in the context of the bus planning process. The TNDP involves planning transit routes and setting frequencies on these routes, in such a way that the resulting transit network would provide the "best" service for a given expenditure of resources. Such planning of the route system is to be accomplished given no a priori specification of a desirable network structure layout. Standard operations research optimization approaches have not been very successful in the solution of the TNDP. The TNDP is combinatorial in nature and presents several sources of nonlinearities and nonconvexities which preclude guaranteed optimal solution algorithms. A solution methodology that relies on AI heuristics and search techniques in addition to domain-specific human knowledge and expertise is presented. It consists of three major components: 1) a route generation design algorithm, 2) an analysis procedure that computes all necessary performance measures and descriptors, particularly on the demand side, and 3) a route improvement algorithm that suggests modifications to the already generated sets of routes. The procedure is implemented in LISP, allowing the convenient application of recursive operations for path search and processing.

We tested the performance of the solution framework by component and as a whole. Our objectives were: 1) to investigate and compare the performance of the solution framework with other proposed solutions to benchmark problems, and 2) to investigate the performance of the solution framework with regard to real world problems. To achieve these objectives we tested the solution framework on the benchmark problem presented by Mandl and on data generated for the transit system of the city of Austin, Texas, U.S.A. In the following sections we overview the three components of the solution algorithm.

The Route Generation Algorithm

The route generation algorithm (RGA) is a design algorithm that configures, for a given transit network and demand matrix, sets of routes that correspond to different trade-offs between the user and operator costs. It queries the user for the minimum percentage of the total demand that has to be satisfied directly (i.e. without transfers) and the percentage of the total demand that is to be satisfied with at most 2 transfers. It searches the demand matrix for high demand node pairs and selects them as seeds for the initial set of skeletons. These skeletons are expanded to routes via different node selection and insertion strategies. The knowledge of transit planners and their expertise are implemented in the different routines in the form of constraints on search and within the different node selection and insertion strategies. Different targets for the demand satisfaction and different insertion strategies result in different sets of routes that require different user and operator costs. RGA relies on algorithmic procedures such as the *k*-shortest paths algorithm, and on the selective application of the transit planners knowledge and expertise to guide the search routines.

Testing of the RGA on data generated for the city of Austin aimed at: 1) investigating the sensitivity of the solution procedure to the initial set of skeletons generated by RGA, and 2) examining the effects of the particular node selection and insertion strategy on the character of the sets of routes generated.

The Analysis Procedure TRUST (Transit Route Analyst)

Once sets of routes are generated, an analysis procedure TRUST is called to evaluate those alternative transit network route configurations. TRUST computes a variety of performance measures reflecting the quality of service and costs experienced by the users, as well as the resources required by the operator. TRUST evaluates a network configuration consisting of a set of transit routes and associated service frequencies. However, bus frequencies are not initially provided to TRUST, because RGA only generates the set of routes and their nodal composition. Thus service frequencies are determined within the TRUST procedure. An initial frequency of service is assumed on all routes, and the given demands between origin-destination pairs are assigned accordingly to the transit network. The frequencies of buses required on all routes to maintain the load factor on all routes under a designer prespecified maximum (which need not be the same across routes) is then computed. If these frequencies are significantly different from the initial assumed value, TRUST reiterates with the output frequencies (i.e. of the preceding run) as input frequencies to the following TRUST execution run. Thus, upon convergence, the output quantities of interest correspond to the frequencies of service necessary and present on all routes.

Route Improvement Algorithm (RIA)

Once the sets of routes are generated by RGA and evaluated with TRUST, the route improvement algorithm (RIA) is called to improve each of the generated sets. RIA is structured in a modular fashion. This allows the planner to assess the impacts of the application of any sequence of modifications (or individual applications of these modules) already considered by TRUST, as well as enables the addition of more and different heuristic improvement technologies, not presently considered by RIA.

Currently, RIA's modifications can be classified into two groups of actions: a) actions on the transit system coverage level, and b) actions on the route structure level. The first goal that RIA was designed to achieve is to make the sets of routes generated by RGA economically and operationally feasible. Testing of RGA on Austin's transit network revealed that as the levels of demand satisfaction approached the 100% mark, the sets of routes generated experienced a considerable increase in the number of routes as well as in total route mileage. A large number of these routes, however, are either too short or suffer from low ridership or both. For such sets of routes, RIA considers two modifications: a) discontinuation of these low ridership routes and/or b) joining these routes or their nodes with other medium-to-high ridership routes.

The second goal of RIA is to demonstrate and test existing improvement procedures suggested by others such as Mandl (1976), Wilson (1982), and Wilson and Gonzalez (1982). The modifications considered are route splitting and branch exchange heuristics (whereby branches of different routes are exchanged to form new routes so as to reduce transfers at the intersection nodes).

In the preliminary testing of RIA, discontinuation of service on low ridership routes was carried out sequentially and in increasing order of their ridership. The two case study sets of routes revealed that route splitting, joining, and discontinuation of service yielded impacts that were consistent with what was expected. Currently, research focusses on investigating the performance of the local branch-exchange heuristics, various sequences of application of the above modifications, and on identifying other possible improvement actions.

Session 9
Exact Methods for the VRP – Méthodes exactes de résolution pour le VRP

Président / Chairperson:	Thomas L. Magnanti
Date:	Friday, June 7 Vendredi, 7 juin
Heure / Time:	11:00
Lieu / Location:	École des H.E.C. Salle / Room 3073

A Cutting Plane Procedure for the Symmetric Generalized Traveling Salesman Problem

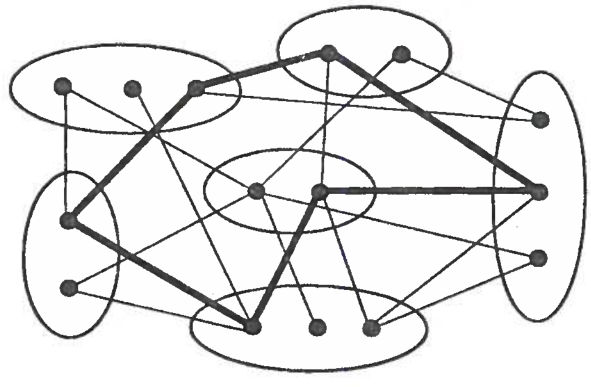
Charles E. Noon and M. Mehdi Sepehri

Management Science Program, The University of Tennessee, Knoxville, TN 37996–0562, U.S.A.

Real-life vehicle routing problems often involve decisions regarding *sequence* and *selection*. *Sequence* refers to the ordering of customer visits. An example of a pure sequencing problem is the Traveling Salesman Problem (TSP). In the context of vehicle routing, the TSP is most appropriate in problems involving a single vehicle operating from a single depot. Most practical problems must consider more than one vehicle or depot and, thus, will necessarily involve an element of *selection*. An example of a pure selection problem is the assignment of customer deliveries to vehicles. The elements of sequence and selection are often addressed separately with one preceding the other. The popular "cluster first-route second" approaches for vehicle routing typify the separation of these two elements.

The focus of our work is a model which combines both elements into a single framework. This model, referred to as the *Generalized Traveling Salesman Problem* (GTSP), allows a decision maker to consider the sequence and selection decisions simultaneously, rather than separately.

Like the symmetric TSP, the symmetric GTSP is defined on an undirected graph G consisting of nodes N, edges $\mathcal{E}$, and edge costs c. Additional specifications are that the nodes are prepartitioned into mutually exclusive *nodesets* and there are no edges defined between nodes of the same nodeset. A feasible *GTSP tour* is a circuit which passes through each nodeset exactly once. The GTSP can be stated as *the problem of finding the minimum cost GTSP tour.* The figure below displays an example GTSP. The nodesets are outlined and a feasible tour is given in bold.



Each node essentially represents a discrete alternative in visiting a nodeset. In vehicle routing, for example, a set of nodes can represent a customer delivery with each node corresponding to a distinct vehicle.

A feasible tour for the GTSP reflects not only a sequence of nodeset visits, but also the selections regarding which nodes are used. If we selected beforehand one node from each set, the problem would reduce to a TSP. On the other hand, if we knew the sequence of nodeset visits, we could use dynamic programming to determine the best selection of nodes to pass through. In the context of vehicle routing problems, the former case is loosely equivalent to a "cluster first-route second" approach with the latter resembling a "route first-cluster second" approach.

The GTSP model provides a substructure that allows us to easily incorporate decision elements beyond those of just selection and sequence. Other elements, such as *balancing* and *timing*, appear in many real problems. Balancing refers to the load assignment decisions when vehicle capacities are present. Timing decisions are required in the case of customer delivery time-windows. The GTSP model serves as a base that can easily accommodate these elements through the introduction of additional constraints and/or variables.

Despite its relevance to vehicle routing, the symmetric GTSP has received little attention in the literature. This is perhaps due to the inherent difficulty in solving its special case, the TSP. The problem was first introduced in 1970 in the context of routing clients through welfare agencies. Early solution approaches were based on dynamic programming but their performance was limited due to state space requirements. In the early eighties, an integer programming based approach for the symmetric GTSP was given by Laporte and Nobert (1983). Their approach relied on a traditional linear programming based branch and bound method.

Our research concerns cutting plane approaches for the symmetric GTSP. Our work is in the spirit of the cutting plane approaches for the symmetric Traveling Salesman Problem (TSP). These approaches use linear programming to solve a relaxed version of the problem and then employ methods for identifying violated constraints. Once violated constraints are identified, they are added to the LP formulation and the problem is resolved. This is repeated until either the LP yields an integer solution or until no more violated constraints can be found. These approaches have been applied to successfully solve rather large symmetric TSP's.

Specifically, our research focus is to investigate several sources of valid inequalities for the symmetric GTSP. Our ultimate goal is to computationally test the effectiveness of these inequalities in a cutting plane procedure. Our work relies on many of the theoretical and computational results for the symmetric TSP. We discuss two ways in which constraint identification methods for the TSP can be used directly for finding violated GTSP constraints. In this sense, our work is an attempt to understand the extent to which the theoretical and computational successes on the symmetric TSP can be applied to real routing problems.

We investigate three sources of valid inequalities for the GTSP and test their effectiveness for LP bounding approaches. Our first source of inequalities is due to a stronger initial formulation for the problem. A previous GTSP formulation given by Laporte and Nobert (1983) yields a somewhat weak LP relaxation since it has no way to enforce that a tour be uninterrupted. Our formulation includes constraints that yield a better initial LP bound on the problem.

Our second source of valid inequalities stems form the fact that any facet inequality for the symmetric TSP can be generalized to a valid constraint for the symmetric GTSP. Violated constraints of this type can be found by identifying violated TSP inequalities over an aggregated solution. The aggregated solution is obtained by summing the interset edge values of any given GTSP solution. This is one way in which we can use the TSP cut identification methods for finding violated GTSP cuts.

The third source of valid inequalities relies on a newly developed problem transformation given in Noon and Bean (1989). These results allow us to efficiently transform any instance of a GTSP into an equivalent asymmetric TSP over the same number of nodes. The transformation provides an important pathway between the GTSP and TSP in that any solution to the GTSP will have a corresponding solution over the equivalent asymmetric TSP. If a transformed solution violates a TSP constraint, the constraint can be inversely transformed to a cut for the GTSP. This allows us to again use the TSP cut identification methods in our search for violated GTSP inequalities. We provide an example that distinguishes this source of inequalities from the two previous sources.

Defining sources of valid inequalities is only half the battle in cutting place approaches. Just as important is the development of a robust strategy for constraint identification and management. We present computational results for the various classes of inequalities and compare the resulting bounds to known optimal objective values.

Significance of the Contribution:

The problem for study is the *Generalized Traveling Salesman Problem* (GTSP). As the name suggests, the GTSP is a generalization of the well known *Traveling Salesman Problem* (TSP). Whereas the TSP is concerned primarily with sequencing, the GTSP is applicable to problems that simultaneously involve *sequencing* and *selection* decisions. It is this added degree of decision making that allows the GTSP to readily serve as a base structure for vehicle routing problems. We investigate a solution approach forth asymmetric GTSP that is based on cutting plane procedures for the TSP. In recent years, these procedures have allowed researchers to solve rather large instances of the TSP. But while the TSP has enjoyed some degree of computational success, there has been little success in solving even medium-sized vehicle routing problems. In an attempt to narrow this technological gap, our solution approach employs a newly developed transformation between the GTSP and the TSP. The transformation provides a pathway that will eventually allow us to channel the theoretical and algorithmic results for the TSP towards solving vehicle routing problems.

Capacitated Trees, Capacitated Routing, and Associated Polyhedra

Jesus Rafael Araque[1], Leslie A. Hall[2] and Thomas L. Magnanti[3]

[1] School of Industrial Engineering, Grisson Hall, Purdue University, West Lafayette, IN 47907, U.S.A.
[2] Department of Civil Engineering and Operations Research, Princeton University, Princeton, NJ 08544, U.S.A.
[3] Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.

We study the polyhedral structure of two related core combinatorial problems: the subtree cardinality-constrained minimal spanning tree problem and the identical customer vehicle routing problem. For each of these problems, and for a forest relaxation of the minimal spanning tree problem, we introduce a number of new valid inequalities and specify conditions for ensuring when these inequalities are facets for the associated integer polyhedra. The inequalities are defined by one of several underlying support graphs:

(i) a multistar, a "star" with a clique replacing the central vertex;
(ii) a clique cluster, a collection of cliques intersecting at a single vertex, or more generally at a "central" clique; and
(iii) a ladybug, consisting of multistar as a head and a clique as a body.

We also consider packing (generalized subtour elimination) constraints, as well as several variants of our basic inequalities, such as partial multistars, whose satellite vertices need not be connected to all of the central vertices. Our development highlights the relationship between the capacitated tree and capacitated forest polytopes and a so-called path-partitioning polytope, and shows how to use monotone polytopes and a set of simple exchange arguments to prove that valid inequalities are facets.

Sensitivity Analysis Methods for Vehicle Routing and Scheduling Problems

Paul M. Thompson* and James B. Orlin**

* Leavey School of Business and Administration, Santa Clara University, Santa Clara, CA 95053, U.S.A.
**Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.

In the operation of many enterprises in both public and private sectors, there often arises the problem of determining cost-effective routes and schedules for a fleet of vehicles. In solving such problems in practice, one is generally not content with merely finding a minimum-cost solution. Rather, one also wishes to determine the sensitivity of the solution and of the solution value to changes in problem data. This is because problem data often are not known with certainty.

For example, when choosing among alternate solutions in a stochastic environment one looks for a robust (i.e., less risky) alternative rather than one with maximum expected gain. Furthermore, rough estimates of costs and other problem parameters are commonly used in practical problem solving because data may be expensive to gather. In this case, it is useful to determine which parameter changes have the greatest effect on solution value, in order to dedicate resources toward precisely estimating the parameters in question, and thereby be able to more accurately estimate true solution costs. Finally, opportunities for improving the current solution often exist in practice because problem constraints may only approximate the real-world situation. For example, "soft" delivery-time constraints may be modeled as "hard" constraints in a model. In such situations, it is useful to know which constraints affect solution quality the most, so that one may check their "firmness", and potentially arrive at an improved solution to the real-world problem.

Sensitivity analysis is the term used for methods that determine the sensitivity of optimal solutions and optimal solution values to parametric variations in problem data. In addition, sensitivity analysis is concerned with finding bounds on coefficient changes that maintain the optimality of the current solution, as well as with systematic methods for reestablishing optimality when these bounds are violated.

Traditionally, sensitivity analysis has focused on measuring the sensitivity of globally optimal solutions to small changes in problem data. Unfortunately, this analysis is not useful with vehicle routing and scheduling problems. The primary reason is that the discrete, non-convex mathematical structure of these problems causes duality gaps, which preclude the use of traditional, dual-based, sensitivity analysis methods. Another important reason is that globally optimal solutions to vehicle routing and scheduling problems are very difficult to obtain. These problems are so complex mathematically that, in practice, one often employs fast algorithms to obtain good, but not necessarily optimal, solutions. Traditional forms of sensitivity analysis do not apply to such approximate solutions of vehicle routing and scheduling problems.

Currently, knowledge of sensitivity analysis methods for approximate solutions to vehicle routing and scheduling problems is severely limited. For example, the recent, comprehensive bibliography of Van Hoesel et al. [1989] on the general topic of sensitivity analysis in combinatorial optimization lists only two papers concerned with the sensitivity of approximate solutions of hard combinatorial problems. Both papers deal with machine scheduling problems.

Because of this meager state of knowledge, we propose and develop a generic framework for sensitivity analysis that applies to vehicle routing and scheduling problems, as well as to other classes of hard combinatorial problems. Our method is based, not on global optimization algorithms, but on approximate algorithms that search local neighborhoods of the current solution. Thus, it is a generalization of traditional linear programming sensitivity analysis, which uses the simplex method to search a neighborhood defined on a polytope. The key difference between the two methods is that LP sensitivity analysis maintains global optimality at all times, while our proposed strategy guarantees only local optimality. Nonetheless, the approaches are quite similar in spirit.

We call the proposed method "local sensitivity analysis", and we develop and discuss several general strategies which fall under this heading. The first is Local Neighborhood Sensitivity Analysis (LNSA). This method measures the sensitivity of the current locally optimal solution value to changes in data, and determines the range of data changes within which the current solution remains a local optimum. Another general strategy is a type of parametric sensitivity analysis, which we call Neighborhood Search Sensitivity Analysis (NSSA). It measures the sensitivity of locally optimal solution value to large data changes, by using neighborhood search to move to new local optima for the modified problem. The neighborhood that NSSA searches is potentially very deep, because the procedure may pass through many solutions in its search for a new local optima. In addition, it becomes even deeper if one uses a look-ahead or multi-step search methodology analogous to the variable depth algorithm of Lin and Kernighan [1973].

We define three general types of problem data: cost data, feasibility data, and structural data; and discuss how local sensitivity analysis methods may be used with them. In general, we find that local sensitivity analysis applies well to cost data, not so well to feasibility data, and not at all to structural data. In particular, local sensitivity analysis applies best to problems that have only cost data, e.g., the TSP.

We develop local sensitivity analysis algorithms using two general neighborhood structures: edge exchanges [Croes, 1958] [Lin, 1965] [Lin & Kernighan, 1973] and cyclic transfers [Thompson & Orlin, 1989] [Thompson & Psaraftis, 1989]. These structures have proved to be useful in finding good approximate solutions to these and other vehicle routing and scheduling problems, and we demonstrate their usefulness with sensitivity analysis as well. In addition, these neighborhood structures are robust (i.e., apply to broad groups of problems), so the likelihood is high that our results will also apply to other problems.

We develop the theory of local sensitivity analysis using these two neighborhoods. First, we consider Local Neighborhood Sensitivity Analysis for edge exchange neighborhoods for the TSP. We show that the upper bound on the range of validity of a basic arc's cost, in a locally optimal TSP solution, is determined by the minimum cost edge exchange that replaces the basic arc with a nonbasic arc. Trivially, the lower bound is infinite. We show that the reverse holds for nonbasic arcs: the upper bound is infinite, while the lower bound is determined by the minimum cost edge exchange.

Then we examine cyclic transfer neighborhoods for the Vehicle Routing and Scheduling Problem with Time Windows (VRSPTW). We derive allowable ranges for the cost data using the auxiliary graph structure of [Thompson & Orlin, 1989] and [Thompson & Psaraftis, 1989], and derive ranges for feasibility data directly from the route structure.

We investigate the effect of multiple simultaneous data changes. In part, this is a feasibility question, i.e. is it feasible to make two or more modifications simultaneously? However, it is also a question of the change in objective function itself. We find that the effect of multiple simultaneous changes may be predicted for both cost and feasibility data if the problem is linear as a function of the data. However, prediction is not sure if the problem is nonlinear.

We show how to use existing data structures to efficiently explore both the edge exchange and cyclic transfer neighborhoods, in order to gather local sensitivity analysis information. Then we build local sensitivity analysis algorithms around these data structures. We derive the computational complexity of these algorithms and, finally, present computational results for both neighborhood structures: edge exchange for the TSP and cyclic transfers for the VRSPTW.

A concluding section discusses the advantages and the limitations of local sensitivity analysis.

References

Croes, G.A. (1958) "A Method for Solving Traveling Salesman Problems", *Oper. Res.* 6, 791–812.

Lin, S. (1965) "Computer Solutions to the Traveling Salesman Problem", *Bell System Tech. J.* 44, 2245–2269.

Lin, S. and B.W. Kernighan (1973) "An Effective Heuristic Algorithm for the Traveling Salesman Problem", *Op. Res.* 21, 498–516.

Thompson, P.M. and J.B. Orlin (1989) "The Theory of Cyclic Transfers", Working Paper OR 200–89, Operations Research Center, MIT, Cambridge, MA, August.

Thompson, P.M. and H.N. Psaraftis (1989) "Cyclic Transfer Algorithms for Multi-Vehicle Routing and Scheduling Problems", Working Paper 89–008, Leavey School of Business and Administration, Santa Clara University, Santa Clara, CA, April.

Van Hoesel, C.P.M., A.W.G. Kolen, A.H.G. Rinnooy Kan and A.P.M. Wagelmans (1989) "Sensitivity Analysis in Combinatorial Optimization", Working Paper, Econometric Institute, Erasmus University Rotterdam, Rotterdam, The Netherlands.

Session 10
Panel

Président / Chairperson:	Warren B. Powell
Date:	Friday, June 7 Vendredi, 7 juin
Heure / Time:	14:00
Lieu / Location:	École des H.E.C. Salle / Room 3073

Panel on Software Development and Implementation

Lawrence Bodin[1], Robert Dial[2], Michael Florian[3], H. Donald Ratliff[4] and Jean-Marc Rousseau[3]

[1] College of Business and Management, University of Maryland, College Park, MD 20742, U.S.A.
[2] Roadnet Technologies, 2311 York Road, Timonium, MD 21093, U.S.A.
[3] Centre de recherche sur les transports, Université de Montréal, C.P. 6128, succursale A, Montréal, CANADA H3C 3J7
[4] Department of Industrial and Systems Engineering, Georgia Institute of Technology, 765 Ferst Drive, Atlanta, GA 30332, U.S.A.

The panel discussion will focus on the following topics:

I. Tales from the trenches - implementing software in the real world

II. The transition from research software to a commercial system
  - Software engineering
  - Robustness
  - Ease of use
  - Maintainability
  - Generality
  - Data handling capabilities
  - Graphical user interface

III. Managing software projects
  - Scoping the problem
  - Identifying the appropriate technology
  - Getting the people

IV. Starting a company
  - Staying on the inside
  - Moving to the outside

V. Software in the corporate world
  - Getting management support for a project
  - Project management
  - Motivating the users
  - System support and maintenance

VI. Future challenges
  - On-line communication with databases
  - Multiple user environments
  - Maintenance and support for customized systems
  - Developing a workforce

Session 11
Stochastic Models II – Modèles stochastiques II

Président / Chairperson:	Dimitri Bertsimas
Date:	Friday, June 7 Vendredi, 7 juin
Heure / Time:	15:15
Lieu / Location:	École des H.E.C. Salle / Room 3072

A Stochastic Programming Approach to O-D Matrix Estimation

Stein W. Wallace* and Kurt Jörnsten**

* Division of Economics, The Norwegian Institute of Technology, The University of Trondheim, N-7034 Trondheim, NORWAY
**Institute for Finance and Management Science, The Norwegian School of Economics and Business Administration, N-5035 Bergen-Sandviken, NORWAY

The purpose of this ongoing research is to find approaches in O-D matrix estimation that do not require consistency in the traffic counts. We argue that, modelling-wise, it is not sound practice to require consistency, as it is not even possible to advocate a scheme for traffic counts that would guarantee consistency. Methods for O-D matrix estimations should therefore view inconsistent traffic counts as a natural part of the input data.

For each link i in a network, let $v_i(k)$, for $k = 1, \ldots, n_i$ be a set of counts. These are different counts made at different times, and hence, it is not considered to be a problem that they are different. Unless there is a reason to believe that some counts are better than others, let us attach the probability $q_i(k) = 1/n_i$ to the count $v_i(k)$. If $Ev_i = \sum_k q_i(k)\, v_i(k)$, we do not even require that these averages are consistent. Or in other words, if p_{ij} is the proportion of travellers in O-D pair j that uses link i, we do not assume that there exists an O-D matrix $\{t_j\}$ such that

$$\sum_j p_{ij}\, t_j = Ev_i \text{ for all } i \tag{1}$$

is consistent. But even so, some version of (1) must be a part of the model.

What we do is the following. Let A be the set of all links for which there are counts. Then search for an O-D matrix $\{t_j\}$ which minimizes the following term

$$Q(t) = \sum_{i\in A} \sum_{k=1}^{n_i} q_i(k) \left(d_i^+ \left(v_i(k) - \sum_j p_{ij}\, t_j \right)_+^{\delta} + d_j^- \left(\sum_j p_{ij}\, t_j - v_i(k) \right)_+^{\delta} \right) \tag{2}$$

where $(x)_+$ equals x if x is positive, and zero otherwise. Note that $d_i^+ \geq 0$ is a penalty for $v_i(k)$ being larger than $\sum_j p_{ij}\, t_j$, and $d_i^- \geq 0$ a penalty for it being smaller. The parameter δ is used to decide how seriously we want to look at outliers.

Problem (2) is a two-stage stochastic program with simple recourse. If $\delta = 1$, $Q(t)$ is piecewise linear. In this case (2) is simple to solve. If $\delta = 2$ then $Q(t)$ is piecewise quadratic.

In most approaches on O-D matrix estimation, one assumes that (1) is consistent, and continue to work with the expectations rather than the original counts. Among all those O-D matrices that satisfy (1) one looks for the most probable or most likely. This is done by means of some function $G(t)$ which could be for example an entropy function. One then minimizes $G(t)$ with respect to (1). This approach can be used in our framework in two different ways. One can either choose to minimize $Q(t)$ to obtain some value Q^*, and then solve

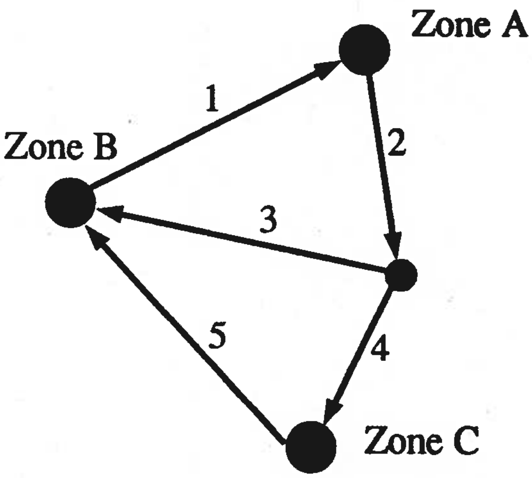
$$\min_{t\geq 0} \{G(t)|Q(t) = Q^*\}$$

or solve immediately

$$\min_{t\geq 0} \{\gamma G(t) + Q(t)\}$$

for some parameter $\gamma \geq 0$. The major idea of our approach, however, namely to view inconsistent counts as a natural part of the input data, is not affected by which formulation one chooses. Note that finding Q^* is our version of "making the counts consistent".

The present state of the work is that we have formulated the model as given above, and applied it to a small problem from Bell (1983). He looks at the following small network.



For each of the five links he assumes the knowledge of five counts. In Bell's context, these counts are five sets of consistent counts. We simply assume that these are a total of 25 independent counts, five on each link. For $\delta = 1$, $q_i(k) = \frac{1}{5}$ and $d_i^{\pm} = 1$ we minimized $Q(t)$ and got that $Q^* = 63$. However, the total flow was far from determined. In fact, one can obtain $Q^* = 63$ with the total flow T varying between 36.7 and 49.3.

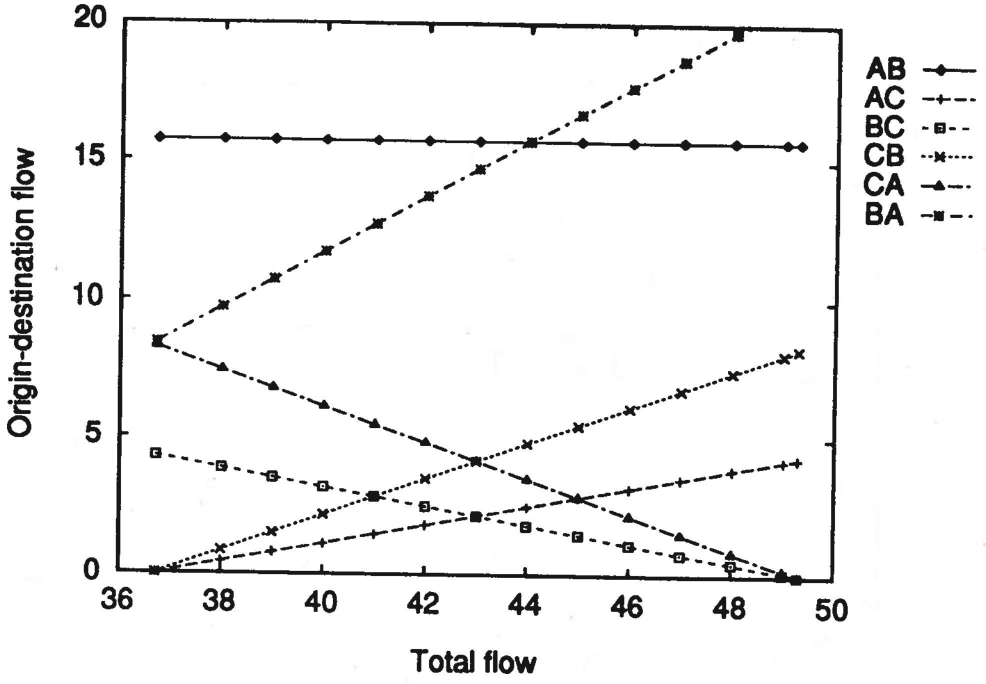
We therefore defined the entropy function $G(t)$ by

$$G(t) = \sum_i t_i \log t_i$$

and then solved the parametric program

$$\min_{t\geq 0} \left\{ G(t)|Q(t) = Q^*, \sum_i t_i' = T \right\}$$

for $T \in [36.7, 49.3]$. The results are shown in the following figure.



The figure shows how the O-D flows depend on the total flow. A major topic for our research until TRISTAN 1 will be to understand this problem of total flow better. In particular we shall compare our approach with other methods to see how they resolve the problem. We also tried the example for $\delta = 2$, and in that case the uncertainty in T was even larger.

Furthermore, we shall try to see how the results from one calculation of an O-D matrix can help us in determining where it would be most useful to obtain a new count. In particular, we would be interested in counts that minimizes the uncertainty in the total flow.

Reference

BELL, M.G.H., "The Estimation of an Origin-Destination Matrix from Traffic Counts", *Transportation Science* 17, 198–217 (1983).

A Stochastic, Dynamic Airline Network Equilibrium Model

François Soumis* and Anna Nagurney**

* GERAD, École des H.E.C., 5255 ave Decelles, Montreal, CANADA H3T 1V6, and Ecole Polytechnique, C.P. 6079, succursale A, Montreal, CANADA H3P 3J7

**Department of General Business and Finance, School of Management, University of Massachusetts, Amherst, MA 01003, U.S.A.

In this paper we develop a stochastic, dynamic network equilibrium model of airline passenger transportation. The model explicitly incorporates the behavior of the passengers in regard to the decision whether to travel or not, and route selection, and formalizes the passenger's decision process as an abstract network. The equilibrium between the realized demand for the routes of travel and the supply of the seats is shown to satisfy a system of nonlinear equations. A Gauss-Seidel algorithm is then proposed for the computation of the equilibrium and conditions for convergence established. Finally, the algorithm is applied for the computation of the Air Canada airline passenger network equilibrium problem.

Reliability in Urban Transportation

Brunilde Sansó* and François Soumis**

* Centre de recherche sur les transports, Université de Montréal, C.P. 6128,
succursale A, Montreal, CANADA H3C 3J7
**GERAD, École des H.E.C., 5255 ave Decelles, Montreal, CANADA H3T 1V6, and Ecole
Polytechnique, C.P. 6079, succursale A, Montreal, CANADA H3P 3J7

Reliability is an important parameter to study in transportation networks. In fact, a reliability study may identify the weak points of the network and could lead to better network planning. However, reliability has not been a parameter widely studied in urban transportation networks. This surprising fact is due in part to the difficulty of finding a good reliability measure for such networks. A traditional approach in Network Reliability has been to evaluate a connectivity measure. This type of measure is mainly related to the network structure and does not necessarily reflect the network capacity to transmit flow, which is the ultimate goal of any flow network. Therefore, connectivity measures cannot be seen as suitable measures for evaluating urban transportation network performance. Moreover, most connectivity measures has been proven or suspected to give rise to NP-hard problems (see, for instance, Ball (1979)). Thus, a connectivity measure would be very hard to calculate in a real-size urban transportation network.

Despite the difficulty in finding a suitable reliability measure and evaluating the measure in real-size networks, some authors have tried different approaches to treat the problem (see, for instance, Hagstrom (1983), Jordan and Turnquist (1979), Turnquist and Bowman (1980) and Sansó and Soumis (1990)).

Sansó and Soumis proposed a general framework to evaluate a measure of reliability in several types of flow networks. They stressed the fact that a good reliability measure in a flow network must reflect the capacity of the network to transmit flow, and the ability of the network to readjust its flow after a failure. The notion behind Sansó and Soumis' approach was that reliability in flow networks depends not only on network structure, but also on the routing decisions used to transmit the flow. The methodology to compute a good performance measure, consisted of enumerating the "key" states of the system, affecting the traffic for every state considered, i.e. solving a routing problem, and evaluating the contribution of those states to the measure chosen.

Transportation networks have the particular characteristic that the accidental state is dynamic. In fact, in many networks, such as telecommunications or electric power, a failure induces an almost instantaneous response on the part of the system, whereas in transportation networks, the flow will have to readjust dynamically to the failure situation, in this case, the accident. In this paper, we tackle the problem of evaluating a reliability measure in a real-size urban transportation system based on the general methodology proposed by Sansó and Soumis (1990) but considering the dynamic aspect of the failure states.

As pointed out previously, the core of the methodology for reliability evaluation of a urban transportation network is the traffic assignment outcome. There are two problems to study in this particular point: one problem is how the user behaves under the knowledge of a possibility of accident in the network. This has been covered in part by the literature on stochastic equilibrium models (see, for instance, Dial (1977), Sheffi and Powell (1981,1982), Mirchandani and Soroush (1987) and the references cited therein). The other problem is how the traffic will be assigned on the routes during the temporal period of the reliability study. In fact, when an accident occurs, a dynamic process starts: some users will be stuck in the arc of the accident, others will see their traveling time increased by the congestion produced by the accident and some will be informed in time that an accident just occurred and will change their usual route. The core of the reliability calculation should therefore be a dynamic traffic assignment model.

Since dynamic traffic assignment packages are still not widely available, we will attempt to solve the dynamic problem approximately. A first model will be created to reproduce the traffic assignment during the period of time in which the reliability study will be done. This period covers the time before the accident occurs, immediately after the accident, and some time after the accident. The main advantage of the model, which implies the resolution of several static traffic assignment problems for every failure state considered, is the availability of traffic assignment packages. We expect to show numerical reliability results and comparisons for a real-size test network.

References

Ball, M.O. (1979) ''Computing Network Reliability'', *Operations Research* 27(4), 823-838.

Dial, R.B. (1971) ''A Probabilistic Multipath Traffic Assignment Model Which Obviates Path Enumeration'', *Transportation Research* 5, 83-111.

Hagstrom, J.N. (1983) ''Using the Decomposition-Tree of a Network in Reliability Computation'', *IEEE Transactions on Reliability* 32, 71-78.

Jordan, W.C. and M.A. Turnquist (1979) ''Zone Scheduling of Bus Routes to Improve Service Reliability'', *Transportation Science* 13(3), 242-268.

Mirchandani, P. and H. Soroush (1987) ''Generalized Traffic Equilibrium With Probabilistic Travel Times and Perceptions'', *Transportation Science* 21(3), 133-151.

Sansó, B. and F. Soumis (1990), ''Communication and Transportation Networks Reliability Using Routing Models'', Publication #548, Centre de recherche sur les transports, Université de Montréal. Scheduled to appear in *IEEE Transactions on Reliability* , April 1991.

Sheffi, Y. and W. Powell (1981) ''A Comparison of Stochastic and Deterministic Traffic Assignment Over Congested Networks'', *Transportation Research* 15B, 53-64.

Sheffi, Y. and W. Powell (1982) ''An Algorithm for the Equilibrium Assignment Problem with Random Link Times'', *Networks* 12, 191-207.

Turnquist, M.A. and L.A. Bowman (1980) ''The Effect of Network Structure on Reliability of Transit Service'', *Transportation Research* 14B, 79-86.

Session 12
VRP with Time Windows –
Routes avec fenêtres de temps

Président / Chairperson:	Jacques Desrosiers
Date:	Friday, June 7 Vendredi, 7 juin
Heure / Time:	15:15
Lieu / Location:	École des H.E.C. Salle / Room 3074

A Variable Depth Approach for the Single-Vehicle Pickup and Delivery Problem with Time Windows

L.J.J. van der Bruggen, J.K. Lenstra, and P.C. Schuur

Eindhoven University of Technology, Department of Mathematics and Computing Science,
P.O. Box 513, 5600 MB Eindhoven, THE NETHERLANDS

In the *single-vehicle pickup and delivery problem with time windows* (SVPDPTW) we are given a single depot, a vehicle with known capacity, and N customers with known demands. Each customer must be picked up at his origin and delivered at his destination and has two time windows in which the pickup and delivery must take place. The problem is to determine a route and a schedule for the vehicle. A route is the ordering of the pickups and deliveries of the customers; a schedule specifies the times of pickup and delivery. Each route starts and ends at the depot. The departure and the arrival time at the depot also have to fall within given time windows. The objective is to minimize the route duration, i.e., the difference between the arrival and departure time at the depot.

The SVPDPTW is a constrained version of the *traveling salesman problem* (TSP) defined on $n = 2N + 2$ vertices. Here the depot is represented by 2 vertices with zero distance. The constraints are imposed by the precedence relations between the origin and the destination of each customer and between the starting and arrival point of the route, as well as by the capacity restriction and the time windows.

In this paper we present a local search method for the SVPDPTW based on a variable depth approach, which can be seen as a generalization of the well-known technique developed by Lin and Kernighan for the travelling salesman problem.

Let us recall that a local search method iteratively tries to improve a solution by local modifications. For the TSP a well-known modification mechanism is a *k-exchange*, i.e., the replacement of k arcs of a given route by another set of k arcs. A route that cannot be improved by a *k-exchange* is said to be *k-optimal.*

As the TSP has no side constraints, the resulting route after a *k-exchange* is always feasible. To check profitability of an exchange, it suffices to compute the difference in cost between the removed and the new arcs. This takes constant time. In case we have n vertices, there are $\Theta(n^k)$ possible k-exchanges so testing a given route for *k-optimality* takes $\mathcal{O}(n^k)$ time. The total time requirement to check a given route for *k-optimality* thus increases rapidly with k. Therefore, one often restricts the search to the cases $k = 2$ and $k = 3$.

In case of the SVPDPTW, one also has to perform feasiblity checks in addition to profitability checks as the resulting route after a *k-exchange* may be infeasible. Checking feasibility and profitability of a single 2–exchange in a straightforward way will take $\mathcal{O}(n)$ time, thus increasing the total time requirement to check a given route for 2–optimality to $\mathcal{O}(n^3)$. Remarkably enough, this total time requirement can be reduced to $\mathcal{O}(n^2)$, i.e., the same time complexity as for the TSP. The idea is to use a search strategy that examines all the 2–exchanges (feasible or not) of a given route in lexicographic order. Moreover, a set of global variables is introduced such that checking feasibility and profitability of a single 2–exchange only take constant time.

It is obvious that the quality of a solution that is ***k-optimal*** improves as *k* increases. We have to find a compromise between a reasonable amount of computing time, requiring a small value of *k*, and the quality of the solution, requiring a large value of *k*. It is therefore, a serious drawback to have to specify the value of *k* in advance. Lin and Kernighan developed a ***variable depth*** exchange procedure for the TSP, i.e., an arc-exchange procedure in which the number of arcs to be replaced is determined dynamically. Given that it has been decided to exchange *s* arcs, heuristic rules are applied to determine whether an *s+1* exchange should be considered.

This variable depth exchange procedure is not applicable to the SVPDPTW in a straightfoward way. Problems concerning feasibility and profitability arise. In this paper we try to overcome these problems by considering the exhange of two arcs at a time instead of swapping one arc as is done by Lin and Kernighan. In each iteration we generate a sequence of 2–exchanges in the following way. In step *s* of the iteration we start from the far most arc added so far and search forward in lexicographic order, using the global variables, for a feasible 2–exchange that makes the total profit obtained by the *s* successive 2–exchanges positive. Thus, in addition to profitable 2–exhanges we also allow deteriorations in cost in a limited way during an iteration. At the end of each iteration the sequence of 2–exchanges that produces maximal profit is chosen. The resulting variable depth exchange algorithm for the SVPDPTW produces solutions that are at least 2–optimal.

This variable depth exchange algorithm considers feasible solutions only. It may seem easy to construct a feasible solution but this is not true. The problem of determining whether there exists a feasible solution to the SVPDPTW is NP-complete.

To obtain a good feasible solution we introduce a two-phase method. In the first phase an initial feasible solution is constructed; during the second phase this solution is improved, using the variable depth exchange algorithm described above.

To construct a feasible route in the first phase, we modify the variable depth exchange algorithm in the following way. In addition to feasible routes we also consider routes that do not satisfy the time constraints. This means that during the first phase routes only have to satisfy capacity and precedence constraints. Note that it is easy to generate such a route. Instead of the route duration, our objective function now measures the total time by which the time windows along the route are exceeded and the search proceeds as long as infeasibility occurs. Unfortunately, it is unlikey that this time violation can be computed in constant time per exchange. However, it turns out that in practical cases, by applying a simple clustering technique on the time windows of the customers, the total computational effort can be reduced considerably.

We have tested this two-phase algorithm on real-life test problems as well as on self-constructed problems. For the real-life problems, the two-phase method produced near optimal solutions. When tested on the self-constructed problems, it sometimes delivered bad solutions.

These bad results can be explained as follows. In either phase the algorithm may get stuck in a poor local optimum. Furthermore, the arc-exchange mechanism we employed does not cover the entire solution space (consisting of all feasible solutions). This means that, given an initial feasible solution, it may not be possible to reach a global optimum. The choice of the initial feasible solution is of crucial importance.

To overcome this problem we have developed an algorithm based on simulated annealing. During the search we consider solutions that satisfy the capacity and precedence constraints, but not necessarily the time window constraints. We choose for a very simple modification mechanism namely the swapping of 2 adjacent vertices. Now we employ an objective function that simultaneously measures the route duration as well as punishes the time window violation. By punishing the time window violation more severely as the search proceeds, one may achieve that the optimization process ends at a feasible solution of high quality. Note that in this way we have integrated the construction phase into the improvement phase. An advantage of simulated annealing opposed to the variable depth exchange algorithm is its flexibility. Annealing is very suited to avoid (bad) local optima; moreover the annealing algorithm will converge asymptotically to an optimal solution. A disadvantage of simulated annealing is its large time requirement.

The latter algorithm produced high quality solutions for all test problems, also in the cases in which the two-phase method did not succeed.

Vehicle Routing with Time Windows: Identification of Problem Structure

Karsten Halse

IMSOR, The Institute of Mathematical Statistics and Operations Research, The Technical
University of Denmark, DK-2800 Lyngby, DENMARK

The Vehicle Routing Problem with Time Windows (VRPTW) consists of finding the routes for a fleet of capacitated trucks, so the cost of the routes are minimized, the customers are assigned to vehicles so the capacities are not exceeded, and they are visited in a specified time period.

It is well known that the VRPTW is NP-hard, and therefore difficult to solve. Until recently there were only heuristic solution procedures to the VRPTW. Though some exact solution procedures have been presented. Kolen et al. have generalized the dynamic programming state space relaxation approach to the VRPTW. Combined with a branch & bound scheme they have solved a 15–customer problem. Knudsen have solved a 30–customer problem with a set partitioning and column generation method. Madsen have chosen a variable splitting approach, and have solved a 30–customer problem. Combined with a branch & bound scheme he was able to solve a 31–customer problem. Desrochers et al. have used a set partitioning and column generation method. With this method they were able to solve problems with 100 customers. They have also proposed a branch & bound scheme to close the duality gap. In general the different methods might perform well on problems with one structure and poorly on problems with another structure. For instance the Set Partitioning approach appears to be particularly efficient for strongly constrained problems.

We will define a measure of restrictiveness of respectively the capacity constraints, the time window constraints, and the interaction between the two types of constraints. A usual measure of the restrictiveness of the time window constraints is the tightness of the time windows, which can be calculated in several ways. The identification of the problem structure and the determination of the restrictiveness serve two purposes.

First, we should be able to, on the basis of the tests on the actual problem, choose the most efficient solution procedure to the actual problem. Second, most of the exact solution procedures presented use a branch & bound scheme for closing the possible duality gap. If we in advance, by identification of the problem structure, are able to tell why the duality gap occurs, we can use this information in the design of the separation rule in the branch & bound scheme. This will evidently reduce the computer time used.

Some computational results with respect to the variable splitting method are reported.

A Two-Phase Heuristic for Vehicle Problems with Time Window Constraints

Mohamed Haouari, and Pierre Dejax

Laboratoire Économique, Industriel et Social, Ecole Centrale Paris, 92295 Châtenay Malabry Cedex, FRANCE

We present a new heuristic algorithm for solving the VRPTW. This method has allowed us to solve a wide variety of 56 benchmark problems efficiently. Thus in 77% of the cases we have improved the best solutions previously published in the scientific literature. The efficiency of our algorithm is due on the one hand to the combination of a new construction method with an improvement method, and on the other to the introduction of a new local improvement strategy that allows the use of several types of neighborhoods to avoid being trapped by local optima. We present a comparison of the performance of our algorithm with several other heuristic methods on the same test problems depending on factors such as time window characteristics, vehicle capacity constraints and the clients distribution.

Session 13
Dynamic Shortest Paths –
Plus courts chemins dynamiques

Président / Chairperson:	Robert Dial
Date:	Friday, June 7 Vendredi, 7 juin
Heure / Time:	16:45
Lieu / Location:	École des H.E.C. Salle / Room 3072

Minimum Time Paths in Dynamic Networks with Application to Intelligent Vehicle/Highway Systems

David E. Kaufman and Robert L. Smith

Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, MI 48109, U.S.A.

An element of the emerging field of Intelligent Vehicle/Highway Systems (IVHS) is the enhancement of travel quality by improving drivers' route choice. The traffic network of freeway corridors and surface streets can be modeled as a directed graph, with some generalized travel cost associated with each arc of the graph, corresponding to a link in the network. Ideally, IVHS would monitor the network through surveillance technologies, measuring these travel costs in real time and disseminating them to hardware installed in individual vehicles. Then all that would remain is for the driver to program his destination (and perhaps his current location, if Automatic Vehicle Location technology lags) and his personal guidance unit would employ a shortest-path algorithm to provide him with an optimal, i.e., least-cost route.

This approach overlooks the rapid change which often characterizes road networks, particularly in rush-hour conditions when benefits of IVHS may be greatest. Changing traffic volumes and incidents causing lane blockages will alter the time required to travel links of the network, and this travel time will generally be a primary, if not the only, component of travel cost. One way to approach this dynamic character of traffic networks is to update the system measurements very frequently, so that the inputs to the shortest-path algorithms are as current as possible. However, this does not address the inherent inaccuracy of the model. Since travel costs will dynamically change, the costs should be modelled as varying with time. (Of course, this creates a related problem, not addressed in this paper, of forecasting these costs for each link of the network over an appropriate horizon.) In this paper we address this need for the case when cost consists solely of travel time. We will provide a rigorous foundation for the modelling and solution of these networks, and we will demonstrate potential benefits to drivers who have the opportunity to determine fastest paths given the dynamic network model. Although discussion centers on the context of IVHS, the results presented are general to networks with costs known to vary over time.

The fastest path problem with time-dependent travel times was first considered by Cooke and Halsey (1966). They cited Bellman's Principle of Optimality to write a functional equation which implicitly defined a network having states incorporating not only the current location in the network (i.e., node), but also the current time. Travel times are assumed to be known for each arc at times $t_0, t_0 + \Delta, t_0 + 2\Delta, \ldots$ and are multiples of Δ, for some $\Delta > 0$. Their solution algorithm proceeds recursively on the maximum number of nodes visited in a path, and if some upper bound $T_{max}\,\Delta$ can be found for the optimal trip time, then the computational effort is $0(N2T_{max}2)$.

Dreyfus (1969) observed that Cooke and Halsey's implicit expansion of the state space and restriction to discrete time intervals can be avoided, and the problem solved by a generalization of Dijkstra's method (Dijkstra (1959)), as efficiently as for static shortest path problems (constant link travel times). However, we will provide a counterexample, and then rigorously establish conditions under which generalizations of conventional static shortest-path algorithms may be applied. We will also discuss how Cooke and Halsey's algorithm may be accelerated by the application of a generalized static shortest-path algorithm when the latter is not guaranteed to provide an optimal solution.

Hall (1977) considers networks with random time-dependent travel times and demonstrates that in general, static algorithms cannot be applied. He notes that adaptive routing is necessary, since policies in which the path chosen from an intermediate node depends on the time of arrival at that node may have expected travel times less than that of any fixed routing policy. We will furnish conditions which allow these networks to be solved by static algorithms at computational cost similar to that of static deterministic shortest path solutions. Because these conditions are somewhat restrictive, we will discuss how static methods can be used as heuristics to improve performance of the general-case algorithm proposed by Hall.

Finally, we will present computational results demonstrating the travel-time reduction which can be realized by a vehicle selecting its route in a road network by anticipatory fastest-path calculation. These results are the product of modifications to the INTEGRATION traffic simulation program (Van Aerde and Yagar (1988)), which is ideally suited to investigation of varying methods of route choice for individual drivers.

Dynamic Shortest Paths with Markovian Arc Costs

Harilaos N. Psaraftis* and John N. Tsitsiklis**

* Department of Naval Architecture and Marine Engineering, National Technical University of Athens, Athens 10682, GREECE.

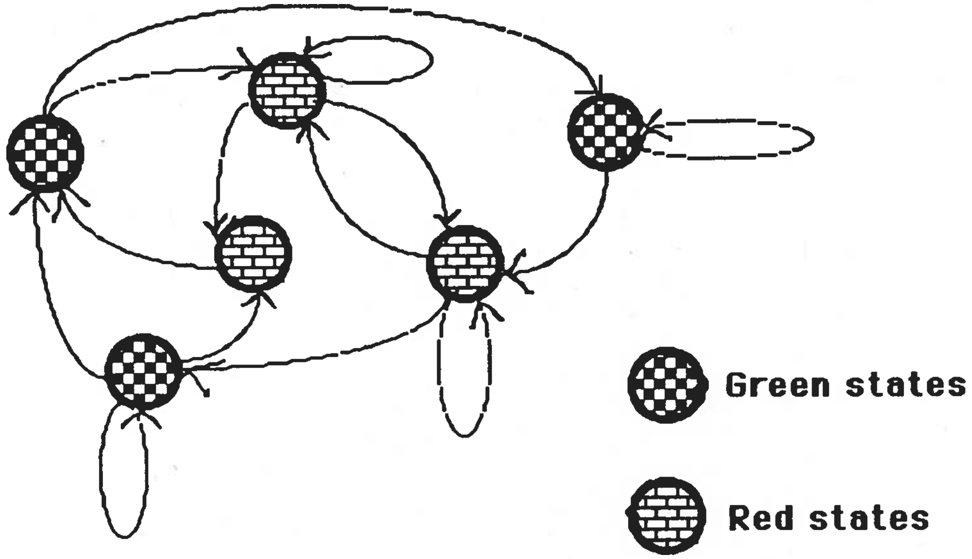
**Laboratory for Information and Decision Systems, and Operations Research Center, Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.

In this paper we examine shortest path problems in a stochastic and dynamic setting, as follows. Assume a directed acyclic graph $G(N, A)$. A vehicle plans to traverse G, starting from a specified node 1, and ending at another specified node n. Assume also that arc traversal costs on G are stochastic and dynamic, in the following sense. The cost of traversing each arc (i, j) of G is a known function $f_{ij}(e_i)$ of the state e_i of a certain "environment variable" at node i at the time the vehicle departs from node i on its way to node j. Environment variables are mutually independent, and each is governed by a finite-state Markov process of a known transition probability matrix. State transitions occur in discrete time. The actual state e_i of the environment variable at node i is revealed to the vehicle only when it is at node i (and if, in fact, it chooses to visit that node). Once at node i, the vehicle may either immediately depart from i toward some other node j and incur the cost $f_{ij}(e_i)$ associated with the prevailing state e_i at departure time, or, it may choose to wait, in anticipation of a "more favorable" environment state at i. Waiting at a node may last as long as desired, but it will cost the vehicle an amount of C per state transition. Based on the above, what is the policy that minimizes the expected total cost of a traversal from node 1 to node n?

Motivation about this problem can come from a variety of transportation contexts, generally involving the motion of a vehicle across a stochastic terrain. One real-world problem that comes close to the abstract version of the problem concerns the routing of a ship sailing across the ocean under uncertain and dynamically changing weather conditions. In this case, the nodes of the network can be thought of as geographical regions across which the ship will transit. The environment variable at each region is a vector of meteorological variables (significant wave height, predominant wave direction, wind magnitude, wind direction, etc.), that describe the state of the weather in that region. Weather conditions change dynamically and stochastically in time. The cost of moving from a certain region to an adjacent region mainly consists of the cost of fuel consumed in doing so, and this is a function of the prevailing weather conditions. If this cost is high enough (due to adverse weather, a storm, etc.), it may make sense for the ship to wait (at a cost), in anticipation of more favorable weather later on.

Other contexts may arise in the trajectory planning of aircraft across a region under stochastic and dynamically changing conditions, in the routing of underwater robots, or even in the hypothetical routing of a car through an urban area, the traffic density of which evolves according to a Markov process.

To solve this problem, we first develop two recursive procedures for the individual arc case, one based on successive approximations, and one based on policy iteration. The policy iteration method has the advantage of being able to solve the problem in at most K iterations, where K is the number of Markov states. We also solve the same problem via parametric linear programming. It is seen that the optimal policy essentially classifies the state of the environment variable at a node into two categories: "green" states for which the optimal action is to immediately traverse the arc, and "red" states for which the optimal action is to wait (see figure below). Various properties of the optimal policy are proven.



We then use these concepts as a building block for the entire network. We develop two dynamic programming procedures, one suboptimal and one optimal, that solve the corresponding problem. The suboptimal procedure is naive in that it is based on the individual arc policy (developed earlier) and is shown to have an arbitrarily bad worst-case error. The optimal procedure uses the policy iteration method developed earlier and has a complexity that is shown to be linear in the number of network nodes and cubic in the number of Markov states. We also discuss using parametric linear programming to solve this problem.

We present examples throughout the paper, and discuss possible research extensions.

Does Providing Information to Drivers Reduce Traffic Congestion?

Richard Arnott[1], Andre de Palma[2], and Robin Lindsey[3]

[1] Boston College, Chestnut Hil, MA 02167, U.S.A.
[2] Northwestern University, Evanston, IL 60208, U.S.A.
[3] University of Alberta, Edmonton, Alberta, Canada T6G 2H4

The purpose of this paper is to question the presumption that route guidance and information systems necessarily reduce traffic congestion, and to point out the need to consider the general equilibrium effects of information. A simple model of the morning rush hour is adopted in which commuters choose a departure time and one of two routes to work, the capacities of which are stochastic. While expected travel costs are reduced by perfectly informing all drivers about route capacities, this is not necessarily the case if imperfect information is provided. A heuristic explanation is that, absent tolls, congestion is an uninternalized externality. Information can cause drivers to change their departure times in such a way as to exacerbate congestion.

Session 14
Routing and Scheduling II – Routes et horaires II

Président / Chairperson:	Marius M. Solomon
Date:	Friday, June 7 Vendredi, 7 juin
Heure / Time:	16:45
Lieu / Location:	École des H.E.C. Salle / Room 3074

Heuristic Algorithms for the Multiple Depot Vehicle Scheduling Problem

Mauro Dell'Amico, Matteo Fischetti and Paolo Toth

DEIS, University of Bologna, Viale Risorgimento 2, 40136 Bologna, ITALY

We consider the NP-hard Multiple Depot Vehicle Scheduling Problem, in which a given set of time-tabled trips have to be assigned to vehicles stationed at different depots, so as to minimize the number of vehicles used and the overall operational cost. The problem arises in the management of transportation companies.

In this paper some structural properties of the problem are studied and used to design a new polynomial-time heuristic algorithm which always guarantees the use of the minimum number of vehicles. Several effective refining procedures are also proposed. Extensive computational results on test problems involving up to 500 trips and 4 depots are reported, showing that the new approach always produces very tight approximate solutions in small computing times and outperforms other heuristics from the literature.

Expert Systems and Operations Research: Mutual Benefits for the Routing and Scheduling of Transportation Vehicles

Jean-Yves Potvin and Jean-Marc Rousseau

Centre de recherche sur les transports, Université de Montréal, C.P. 6128, succursale
A, Montréal, Québec, H3C 3J7 CANADA

Artificial intelligence (AI) and Operations Research (OR) techniques are both aimed at supporting decision making. The AI and OR communities should thus mutually benefit from a close cooperation. In fact, more and more scientists are working at the frontiers of AI and OR and build innovative systems that integrate techniques from both domains. The aim of this paper is thus to illustrate this trend in the vehicle routing and scheduling domain, drawing from our own experience.

Arc Routing and Scheduling: New Algorithms and Applications

Michael Ball[1], Lawrence Bodin[2], Hedlye Gun[3], Lawrence Levy[4] and Ivan Su[2]

[1] College of Business and Management and Systems Research Center, University of Maryland, College Park, MD 20742 U.S.A.
[2] College of Business and Management, University of Maryland, College Park, MD 20742 U.S.A.
[3] Applied Mathematics Program, University of Maryland, College Park, MD 20742 U.S.A.
[4] DISTINCT Management Consultants, 10705 Charter Dr., Ste. 440, Columbia, MD 21044 U.S.A.

In this talk we summarize a variety of recent work on arc routing and scheduling problems. Specifically, we describe three different application environments and show how each of these gives rise to slightly different model and algorithmic requirements. Algorithmic work includes the following: investigations into the facial structure of the directed rural postman polyhedron, a dual ascent algorithm for the directed rural postman problem and a heuristic algorithm for the mixed rural postman problem.

We discuss routing and scheduling applications for household refuse pickup, utility meter readers and postal carriers. Important distinguishing features of applications are whether the mode of travel is walking, driving or both, the role capacity plays, the presence of one or more central facilities and various workrules. The combinations of these constraints give rise to either directed, undirected or mixed underlying networks. In all cases, it is possible that the set of arcs that must be served does not induce a connected network so that "rural" postman problems arise. All applications involve the generation of multiple routes so that the output required is both a partition of the required arcs into the individual routes and the sequencing of the arcs withing each route.

All of the application environments mentioned above give rise to complex problems. We discuss multiple step strategies to solving these problems. In addition we show where various "core" simpler problems arise and then describe how optimization algorithms for these problems can be used effectively. One of these core problems is the directed rural postman problem and also give computational results with a dual ascent algorithm for solving it.

Session 15
Computational Methods –
Méthodes numériques

Président / Chairperson:	Judith Farvolden
Date:	Saturday, June 8 Samedi, 8 juin
Heure / Time:	08:30
Lieu / Location:	École des H.E.C. Salle / Room 3073

Parallel Computation of Large-Scale Dynamic Market Network Equilibria via Time Period Decomposition

Anna Nagurney and Dae-Shik Kim

Department of General Business and Finance, School of Management, University of Massachusetts, Amherst, Massachusetts 01003, U.S.A.

The principal issue in the study of dynamic market equilibrium problems is the computation of the optimal commodity production, consumption, trade, and inventory patterns over space and time. Such models are inherently large-scale and, hence, the development of efficient computational procedures is essential for the operationalism of such models.

Dynamic competitive spatial market models have their foundations in the work of Samuelson and Takayama and Judge with a variety of applications, including agricultural and energy markets. Early equilibrium models were reformulated as optimization problems with the observation that in the case of "symmetric" interactions, the equilibrium conditions were, in fact, the Kuhn-Tucker conditions of an appropriately constructed minimization problem. Recently, variational inequalities have been used to formulate more general spatial equilibrium problems.

In this paper we introduce a dynamic market equilibrium model, establish the equilibrium conditions, and then formulate them as a variational inequality problem. We then develop a parallelizable variational inequality decomposition algorithm which takes advantage of the special dynamic network structure of the problem.

The decomposition algorithm decomposes a dynamic market equilibrium problem with T time periods into $T + 1$ subproblems, each of which can be allocated to a distinct processor, and, hence, solved simultaneously and in parallel. The first T subproblems are static spatial price equilibrium problems with special network structure for which numerous efficient algorithms exist, whereas the $T + 1$ subproblem is the inventory problem. Parallel computation of spatial market equilibrium problems, thus far, has focused on decomposition by commodities.

We establish conditions for convergence of the parallel decomposition scheme using variational inequality theory and then present numerical results for the algorithm when it is implemented first in serial fashion and then in parallel.

The numerical results, thus far, suggest that the CPU time increases linearly, rather than quadratically as is typical of other approaches, as the number of time periods is increased. The algorithm also exhibits substantial speedups when it is implemented on a parallel architecture.

Although the focus of our paper is on dynamic spatial market equilibrium problems the results are applicable to dynamic network equilibrium problems in general which exhibit a characteristic structure.

Solution of the Shortest Path Problem on Massively Parallel Computers

Mayiz B. Habbal, Haris N. Koutsopoulos and Steven R. Lerman

Department of Civil Engineering, Massachusetts Institute of Technology, Cambridge,
Massachusetts 02139, U.S.A.

Advances in parallel computer architectures open new opportunities for research on algorithms for the solution of network problems and provide hope that larger and more realistic problems can be solved in a fraction of the time required on serial computers. In this paper we examine a very basic network problem, the shortest path problem, and present approaches for its solution on a novel computer, the Connection Machine.

A large number of parallel computer architectures exist which differ with respect to the type and number of processors they utilize, the way processors are interconnected and communicate with each other, the control mechanism, etc. A popular classification scheme, known as Flynn's taxonomy, distinguishes parallel computers between SIMD and MIMD architectures. SIMD (single instruction/multiple data stream) machines retain a single stream of instructions but support vector operations on multiple data items; at any given time a single operation is at the same state of execution on multiple processing units, each manipulating different data. MIMD (multiple instruction/multiple data stream) computers can execute different instructions on different processing units. Most multiprocessor configurations belong to this class.

The Connection Machine is a SIMD fine-grained, massively parallel computer. It comes in several configurations; the full scale version contains a total of 65,536 bit-serial processors operating at a clock cycle of 250 nsec. Each processor may have up to 32K bytes of memory; interprocessor communication is achieved through a flexible 12–dimensional hypercube communication network. An interesting feature of the Connection Machine is the concept of virtual processors which permits applications to use a higher number of processors than physically available. The use of the virtual processor capability of the machine, however, implies some degradation in performance.

The Connection Machine is programmed in a data parallel style. The parallelism results from simultaneous operations on large data sets, rather than multiple threads of program control. Hence a key element is designing algorithms for the Connection Machine and other massively parallel computers is the mapping of data onto processors to facilitate the operations of the algorithm.

The shortest path problem is a classical and important combinatorial problem with many applications. We examine the solution on the Connection Machine of two versions of the problem: the one-to-all and the all-to-all shortest path.

For the one-to-all problem we discuss the parallel implementation of various existing algorithms and mapping strategies of the network to processors. We also examine non-linear approaches based on Bregman's entropy algorithm, which has been applied successfully to other classes of network problems. The Bellman-Ford algorithm has demonstrated superior performance in all test problems. As expected its running time is proportional to the length (in number of arcs) of the shortest path from the origin to the farthest (in number of arcs) destination. The time required per iteration is almost constant for a wide range of network sizes. The performance of the algorithm does not have any significant advantage when compared to the running time of the best serial algorithm implemented on a VAX computer and is inferior to the performance of the same algorithm on a CRAY computer.

Most of the serial algorithms which have been proposed for the solution of the all-to-all shortest path problem either solve the one-to-all problem for every node or use a matrix representation of the network to determine the shortest paths. Depending on the density of the network one approach may have computational advantages over the other. To solve the all-to-all problem on the Connection Machine (and other massively parallel architectures) we suggest the following network decomposition stragegy:

Phase 0 : Decompose the network into subnetworks and identify the set of cutset nodes associated with each subnetwork.

Phase 1 : For every cutset node determine the shortest path to all other nodes in the network.

Phase 2 : For each subnetwork solve the all-to-all shortest path problem using matrix operations (such as Floyd's trip operation).

Phase 3 : Combine the results from Phases 1 and 2 using matrix operations, to obtain the final shortest path distances from every node to all other nodes.

The algorithm described above takes advantage of the natural representation of matrices on the Connection Machine and its performance depends, among other things, on the following factors:

1. The size of the subnetworks and the number of cutset nodes.
2. The mapping of the network data.
3. The flow of information and data from one phase to the next.

We examine the sensitivity of the execution time of the algorithm to the above parameters and we draw conclusions on appropriate network decomposition strategies. We present efficient mapping schemes and processor organization, and discuss architecture specific issues (such as the use of the virtual processor capability of the Connection Machine). Preliminary results are very encouraging and indicate that the performance of the suggested approach exceeds the performance of the best serial algorithm on both the VAX and CRAY computers.

For all computational experiments reported in the paper we use two transportation networks (a small suburban network with 205 links and 578 arcs and a large urban network with 4,923 links and 11,647 arcs) and grid networks of various sizes (with simulated link costs).

A Splitting Equilibration Algorithm for the Computation of Large-Scale Constrained Matrix Problems: Theoretical Analysis and Applications

Anna Nagurney* and Alexander Eydeland**

* Department of General Business and Finance, School of Management, University of Massachusetts, Amherst, Massachusetts 01003, U.S.A.

**Department of Mathematics and Statistics, University of Massachusetts, Amherst, Massachusetts 01003, U.S.A.

In this paper we introduce a general parallelizable computational method for solving a wide spectrum of *Constrained Matrix* problems. The *Constrained Matrix* problem is a core problem in numerous applications. These include the estimation of origin/destination flows in transportation and telecommunications, the estimation of input-output tables, trade tables, and social/national accounts, the projection of migration flows over space and time, the treatment of census data, and the estimation of contingency tables in statistics. The constrained matrix problem, so named by Bacharach, is to compute the best possible estimate X of an unknown matrix, given some information to constrain the solution set, and requiring either that the matrix X be a minimum distance from a given matrix X^0, or that X be a functional form of another known matrix. In real-world applications, the matrix X is often very large (several hundred to several thousand rows and columns), with the resulting constrained matrix problem larger still (with the number of variables on the order of the square of the number of rows/columns; typically, in the hundreds of thousands to millions). In the classical setting, the row totals and the column totals are known and fixed, and the individual matrix elements nonnegative. However, in certain applications, the row and column totals need not be known a priori, but must be estimated, as well. Furthermore, additional objective and subjective inputs are often incorporated within the model to better represent the application being studied. It is the solution of this broad class of large-scale constrained matrix problems in a timely fashion that we address in this paper.

The constrained matrix approach has become a standard modelling tool among researchers and practitioners. Therefore, the need for a unifying, robust, and efficient computational procedure for solving constrained matrix problems is of practical importance. In this paper we introduce an algorithm, which we call the *splitting equilibration algorithm*, for computing solutions to the entire class of constrained matrix problems. This algorithm is not only theoretically justified, but also fully exploits both the underlying structure of these large-scale problems and the advantages offered by the state-of-the-art computer architectures, while simultaneously enhancing the modelling flexibility.

In this paper we show how the constrained matrix problems can be formulated and studied in a unified fashion as mathematical programming problems. In particular, we introduce a general quadratic constrained matrix problem which allows any positive definite quadratic form as objective function, assumes nonnegative estimates, and, in addition, allows the estimation of the row and column totals. We show that this general model, contains, as special cases, many currently used models.

Our computational procedure is motivated, in part, by the problem at hand – the "equilibration" of matrices (cf. Van der Sluis), and by the problem's connection with spatial price equilibrium problems. In spatial price equilibrium problems, one seeks to compute the commodity production, consumption, and trade patterns, satisfying the equilibrium property that trade takes place between two spatially separated markets if the demand price at the demand market is equal to the supply price of the commodity at the supply market plus the cost of transportation between the pair of supply and demand markets. Indeed, although as early as 1951, Stone recognized that the same methodology should be applied to the computation of both spatial price equilibrium problems and constrained matrix problems, the computational state-of-the-art at the time precluded such an investigation.

In particular, we utilize some recent results from the theory of variational inequalities, to construct a splitting equilibration algorithm which splits a large-scale constrained matrix problem into series of row (supply)/column(demand) equilibrium subproblems. Each such constructed subproblem, due to its special structure, can, in turn, be solved simultaneously via *exact equilibration* in closed form. Thus each subproblem can be allocated to a distinct processor. To date, the theory of variational inequalities has been applied to formulate, study qualitatively, and compute the solution to a plethora of equilibrium problems, including traffic network equilibrium problems, the above mentioned spatial price equilibrium problems, oligopolistic market equilibrium problems, the general economic equilibrium problem, and migration equilibrium problems.

Our approach is in agreement with recent research in which constrained matrix problems are formulated as mathematical programming problems, with an objective function which forces "conservatism" on the process of rationalizing X from the initial estimate X^0. This formalism is helpful and natural for three reasons. Firstly, if viewed from the perspective of mathematical statistics, the quadratic penalty function gives as solution the minimum variance unbiased linear estimate of the matrix X. Secondly, if viewed from the perspective of information theory, the entropy function gives rise to the estimate X which minimizes the "information added" to X^0 needed to conform to the constraints. Lastly, Bacharach has shown that a particular functional form, the result of application of the "RAS" method, is equivalent to constrained entropy minimization. Although the RAS method, which dates to Deming and Stephan in 1940, is currently the most widely applied computational method in practice for the solution of the constrained matrix problem, its limitations include, the use of a highly specific set of contraints and objective function and its nonconvergence in applications, such as interregional trade.

In this paper we first present the formulation of the general quadratic constrained matrix problem, specializing it, firstly, to the particular constrained matrix problem arising in the estimation of social accounting matrices in which the row and column totals must "balance", and, secondly, to the case of known row and column totals. We also relate the general formulation to many currently used models and establish its relationship to spatial price equilibrium problems.

We then develop the splitting equilibration algorithm and then specialize it to particular cases. We then provide a theoretical analysis of the algorithm, including convergence results and computational complexity results. The theoretical analysis is based on the interpretation of the algorithm as a dual method. We also demonstrate how the algorithm can be applied to compute the solution to asymmetric spatial price equilibrium problems, for which no equivalent optimization formulations exist.

We then proceed to empirically investigate the computational performance of the splitting equilibration algorithm on the largest quadratic constrained matrix problems reported to date. Our goals include: 1) to compare the relative efficiency of the splitting equilibration algorithm to several other algorithms, 2) to investigate the efficiency of the new equilibration approach on the spectrum of very large constrained matrix problems, both diagonal and general, and 3) to investigate the speedups obtained with parallelization of the splitting equilibration algorithm for both diagonal and general large-scale problems.

We first consider the case of a diagonal quadratic matrix and present computational results on large-scale problems when the algorithm is implemented, first in serial, and then in parallel, using the full multiprocessor features of the IBM 3090–600E. The datasets include: input/output matrices, social accounting matrices, migration tables, and spatial price equilibrium problems.

We then turn to the computation of general quadratic constrained matrix problems and present computational results in a serial, and then in a parallel, environment.

The Status of Massively Parallel Algorithms for Network Structured Problems

Stavros A. Zenios

Decision Sciences Department, The Wharton School, University of Pennsylvania,
Philadelphia, PA 19104, U.S.A.

We will survey current research on the design of massively parallel algorithms for network structured problems: (i) assignment problems, (ii) linear and nonlinear network flow problems, (iii) multicommodity network flow problems, and (iv) stochastic network programs. In addition to the algorithms we will survey current computational experiences and report on a comparative study of massively parallel algorithms on the Connection Machine CM-2 with coars-grain decomposition methods on the CRAYY-MP.

Session 16
Dynamic Traffic Assignment – Affectation dynamique du trafic

Président / Chairperson:	Anna Nagurney
Date:	Saturday, June 8 Samedi, 8 juin
Heure / Time:	10:45
Lieu / Location:	École des H.E.C. Salle / Room 3073

Algorithms for the Dynamic Network Equilibrium Problem

Terry L. Friesz* and Roger L. Tobin**

* George Mason University, Fairfax, VA 22030, U.S.A.
**GTE Laboratories, Waltham, MA 02254, U.S.A.

Mathematical formulations of the simultaneous route and departure time choice problem on a general network, which is the basis of dynamic network flow predictions, have recently been achieved [e.g. Friesz et al. (1990) and Richter et al. (1990)]. These models allow description of the time histories of flows throughout a network, but only at the expense of either postulating *a priori* trip delay operators for each route and departure time combination or endogenously determining such delays from the underlying queuing processes of network.

There is little basis for assuming in advance specific operators for route/departure combinations, since virtually no empirical investigations of the matter exist. Indeed, it can be demonstrated that certain assumptions about these operators normally made in static problems – most notably, monotonicity – will never be satisfied for realistic dynamic models involving simultaneous route and departure time decisions. Furthermore, as demonstrated by Friesz et al. (1990), the endogenous determination of delay operators involves the difficult (though, by no means, impossible) task of solving a system of simultaneous integral equations which implicitly define those operators.

The model developed by Friesz et al. (1990) is expressible as either an infinite dimensional variational inequality or as an infinite dimensional nonlinear complementarity problem, for which the key delay operators are, as mentioned above, the solution of a system of integral solutions. In fact we may list the following complicating factors with regard to the solution of dynamic equilibrium models expressed as infinite dimensional variational inequalities:

1. Regardless of technique used to solve the variational inequality, one must solve the system of integral equations implicitly defining the delay operations.
2. Path variables are intrinsic to the problem formulation, and it appears that all relevant paths must be enumerated in advance.
3. The variational inequality has no equivalent optimization problem, nor can it be symmetrized as in static problems; hence, it is necessary to use a projection to define an equivalent fixed point problem which may be solved iteratively. This projection involves an infinite dimensional optimization problem which must be solved at each fixed point iteration.

We see from the above discussion that numerical solution of the dynamic network equilibrium problem is potentially very difficult. Our main goal in this paper is to articulate a convergent numerical algorithm for this problem. That algorithm will be based on first transforming the variational inequality to a fixed point problem in Hilbert space by employing a projection. This will involve the repeated solution of an infinite dimensional minimization problem required to set up the projection which makes the fixed point representation possible. A key task will be devising efficient techniques for solving this subproblem, the repeated solution of which will constitute the major workload of the algorithm. The fixed point problem itself will be solved by classical iterative schemes.

As time permits, numerical tests of the new algorithm will be carried out, although detailed numerical experiments to characterize dynamic flow patterns on transport networks will be the subject of a separate paper.

It should be kept in mind that the dynamic equilibrium model motivating this paper is wholly original, not previous reported on from an algorithmic perspective, and behaviorally more realistic than other models reported to date. Especially important is the perspective we take in both the model and the envisioned algorithm that time should be treated in a continuous fashion.

References

Friesz, Smith and Tobin (1990) "A Variational Inequality Formulation of the Dynamic Network Equilibrium Problem", Working paper, Network Analysis Laboratory, George Mason University.

Richter et al. (1990) "Computation of Dynamic User Equilibria in a Model of Peak Period Traffic Congestion with Heterogeneous Commuter", Working paper, Department of Economics, Boston College.

Solving the Dynamic User-Optimal Traffic Assignment Problem Over Congested Transportation Networks

David E. Boyce*, Bin Ran* and Larry J. LeBlanc**

* Urban Transportation Center, University of Illinois at Chicago, 1033 W. Van Buren Street, Chicago, IL 60607, U.S.A.

**Owen Graduate School of Management, Vanderbilt University, 401 21st Ave. South, Nashville, TN 37203, U.S.A.

The development of advanced route guidance systems provides new opportunities to improve traffic conditions and to reduce recurrent and nonrecurrent congestion in urban and suburban transportation networks. In a dynamic route guidance system, the central controller will be informed about every tripmaker in terms of origin, destination and timing of the trip. This information will be used to develop a coordinated strategy which assigns each vehicle to a route to its destination, on a real time basis, so as to achieve the desired user-optimal or system-optimal objectives.

In this paper, the dynamic user-optimal (DUO) traffic assignment problem is studied, which assigns vehicles so that each driver uses the instantaneous minimal-cost path employing the currently prevailing travel times. Using the optimal control theory approach, a new DUO traffic assignment model is formulated for congested transportation networks. This DUO traffic assignment model is a dynamic generalization of the static user-optimal model. By employing link exit flows as another set of control variables, rather than exit functions, this model is fundamentally different from earlier dynamic user-optimal models. The equivalence of this optimal control program with the DUO traffic assignment is demonstrated by proving the equivalence of the first order necessary conditions of the program with the instantaneous DUO conditions. Since this DUO model can provide information about the instantaneous minimal-cost route between each origin-destination pair, it may be very useful for controlling vehicles in a dynamic route guidance system.

In order to solve the DUO traffic assignment efficiently for a large-scale network, the assignment time period is subdivided into small time intervals and the optimal control problem is reduced to a nonlinear programming (NLP) problem. An algorithm based on the Frank-Wolfe technique is proposed to solve this NLP problem. The algorithm's linear subproblem for direction finding is solved by searching the minimal-cost path between each OD-pair for each time interval and assigning instantaneous OD trips at the same time.

The matrices of dynamic traffic assignment problem are typically very large and sparse, which results in a large number of operations and a large amount of memory required for an optimal solution. However, the separability of the DUO problem according to origin-destination pair and time period, which is fully employed in our algorithm, makes this problem a good candidate of parallel and vector processing. For a large-scale problem with more than 1000 links and 10 time periods, a substantial saving of computing time may be expected by using the supercomputing techniques. A moderate-scale test network will illustrate this computation. Computational results on both the IBM mainframe computer and the CRAY supercomputer will be reported.

Stochastic Process Modelling of Transportation Network Dynamics

Ennio Cascetta and Giucco Erberto Cantarella

Dipartimento di Ingegneria dei Trasporti, Università degli Studi di Napoli, Via Claudio 21, 80125 Napoli, ITALY

A stochastic process approach to the modelling of dynamics in transportation networks, recently proposed by the authors, is described and extended. The approach is based on an explicit modelling of the users' day-to-day adjustment process.

This approach allows to overcome some of the major drawbacks of the usual equilibrium analysis (in which a self-reproducing or fixed-point solution is searched) namely the assumptions of existence, uniqueness and stability of the equilibrium needed to assure that the system state converges toward the equilibrium state and that, once reached, it is stable with respect to small fluctuations.

Under mild assumptions this stochastic process model admits one and only one steady state probability distribution, independent of the adopted network performance model (cost-flow functions, etc.). Furthermore, it can take into account real-time dynamic users' behavior, that is while-trip re-routing.

In this paper, the application of the stochastic process approach to the modelling of the day-to-day network dynamics assuming within-day constant travel demand is described first. Then, the approach is extended to include the modelling of the within-day profile of travel demand, by simulating users' departure time and path choices. En-route diversions from the initially chosen path is also modelled.

The practical application of a day-to-day and within-day doubly dynamic model requires a network loading method simulating time-varying link flows and costs resulting from a variable demand.

Algorithmic issues are addressed showing that the proposed approach can be successfully applied to real-size networks. Among others it can be effectively used to estimate the effects of information control strategies, including route guidance systems based on real time information about demand fluctuations and/or incident detection.

An Algorithm for Dynamic Traffic Assignment Models Based on an Optimal Control Approach

Jaime Barceló and E. Codina

Department of Statistics and Operations Research, Universitat Politecnica de Catalunya,
Pau Gargallo, 5, 08028 Barcelona, SPAIN

The dynamics of traffic flows is of special importance to study dynamic behaviours like for instance that of route guidance systems in urban networks. In the last few years, several authors (Ran, Friesz, Boyce, Wie, Papageorgiou, ...) have introduced optimal control theory to model dynamic user equilibrium problems. Although theoretical properties of that formulation have been well established, research of efficient algorithms is far from being closed, especially for the case of large urban networks.

In this paper the problem of Dynamic User Optimal Traffic Assignment using Optimal Control formulation with free terminal conditions is stated and examined in order to outline a solution algorithm that uses as a subproblem the well-known static user traffic equilibrium assignment problem. The described algorithm is based on a time decomposition method and a flow aggregation scheme that makes the resulting mathematical programming formulation quite natural to appear and especially suited to solve large urban network problems.

Because of the free terminal conditions, the subproblems for time subintervals can be solved isolately once the previous subproblem solution is known. In order to improve computational performance two specially designed techniques are described: the first one allows to solve the subproblem at a step without "starting from scratch" and the second one applies when a restricted simplicial decomposition algorithm is used to solve the static user traffic equilibrium assignment subproblem. The efficiency of this technique is discussed against the use of sensitivity analysis and computational results are given. Finally, for the case of large urban networks, the algorithm is matched with a dualization based one and the case of fixed terminal conditions is formulated and a solution algorithm described.

Session 17
VRP Heuristics – Méthodes heuristiques pour le VRP

Président / Chairperson:	Michael O. Ball
Date:	Monday, June 10 Lundi, 10 juin
Heure / Time:	09:00
Lieu / Location:	Château Frontenac Salon Frontenac

A TSSP+1 Decomposition Approach for the Capacity-Constrained VRP

John Mittenthal and Charles Noon

Department of Decision Sciences and Engineering Systems, Rensselaer Polytechnic
Institute, Troy, New York 12180–3590, U.S.A.

This paper concerns an approach for the basic, capacity-constrained vehicle routing problem (VRP). For the basic VRP, m vehicles are stationed at a single depot and must be used to deliver product to a collection of n customer cities. Each customer city i must be supplied with a particular amount of product, $d(i)$, and each vehicle k has a specific limit on the amount of product it can supply, $CAP(k)$. The problem is to determine a set of capacity-feasible vehicle routes with minimum total travel distance. A successful strategy for solving the VRP has been to decompose the problem into a master (or dispatcher level) problem and a collection of subproblems (or driver level) decisions. For example, in the well-known approach of Fisher and Jaikumar (1981, Networks), the dispatcher is responsible for making capacity-feasible customer/vehicle assignments and the drivers are responsible for sequencing their individual routes. In their approach, the master problem is modeled as a generalized assignment problem with the resulting individual subproblems being Traveling Salesman Problems over the assigned customers.

The approach presented in this paper is also based on a problem decomposition which separates the decisions of a dispatcher from those of the individual drivers. What distinguishes our approach is that it places a greater share of the decision making on the individual vehicle driver. In our approach, the dispatcher is responsible for assigning only one customer to each vehicle and then assigning a reward value to each of the remaining customer cities. Based on these reward values together with the given travel costs, each vehicle driver is responsible for choosing which of the unassigned customers to visit and determining an individual capacity-feasible route. The challenge of the dispatcher is to assign reward values such that each customer is visited by exactly one vehicle.

Our approach is founded on a Lagrangian relaxation whereby the relaxed constraints require that each customer be visited by exactly one vehicle. The problem faced by the dispatcher is, in essence, to choose an appropriate set of Lagrange multipliers for the relaxed constraints. The subproblem faced by each driver is modeled as a Travelling Salesman Subset-tour Problem with one additional constraint (TSSP+1). The TSSP+1 problem has as its special case the Orienteering Problem, the Time-constrained TSP and the Prize Collecting TSP.

The purpose of this paper is to investigate the viability of the TSSP+1 decomposition strategy for the basic VRP and to test its computational effectiveness. One contribution of this research is to bring together existing approaches and results for the TSSP+1 class of problems towards solving multi-vehicle routing problems.

Solving Real-Life Vehicle Routing Problems with Tabu Search: Two Adaptations

Frédéric Semet

Département de Mathématiques, École Polytechnique Fédérale de Lausanne, MA (Ecublens), CH-1015 Lausanne, SWITZERLAND

The tabu search metaheuristic has enabled people to obtain very good results for a wide range of combinatorial optimization problems such that the quadratic assignment problem, the travelling salesman problem, the job-shop scheduling problem...

This method has also been very successful to solve real-life problems when many different types of restrictions must be taken into account. For instance, A. Hertz has solved timetabling problems including seven different constraint types. Unfortunately, the best heuristics for solving the VRP cannot deal with such diversified constraints. Therefore, for solving two-real vehicle routing problems, we use the tabu search technique, adapting it in a specific way to each of them.

The first case study concerns the delivery of goods to customers using a heterogeneous fleet which includes trucks and trailers. The different types of constraints which are taken into account are the following:

- weight and volume restrictions
- time windows
- total accessibility restrictions: some customers can be served by some, but not necessarily all vehicles
- partial accessibility restrictions: some customers can be served only by trucks, other by either trucks or trailers. Obviously, in our solution, we have to allow routes which include these two kinds of customers.

The main characteristics of the tabu search heuristic developed for this problem is the conservation of an admissible solution at each step. The algorithm can be summed up as follows: initially, consider routes which consist of one customer and then, at each iteration, modify the routes, and if they are admissible, assign them to the trucks.

With this method, we solve different instances of the problem involving from 50 to 100 customers, and we compare our results to the current solutions used by the company.

The second real-life problem deals with the delivery of animal food to farmers. The fleet used is homogeneous and the different types of constraints are the following:

- capacity restrictions
- total time restrictions
- time windows
- total accessibility restrictions.

We have a different approach to solve this problem: first, an admissible solution is computed with a classical insertion heuristic, then a tabu search procedure is used in which some constraints are relaxed.

The different instances of this problem involving from 100 to 150 customers are solved in order to compare with the current solutions used by the company.

Finally, these two different approaches are discussed relatively to our experience and some concluding remarks are given.

A Tabu Search Heuristic for Vehicle Routing Problems

Alain Hertz*, Michel Gendreau and Gilbert Laporte****

* GERAD and Ecole des Hautes Etudes Commerciales, 5255, avenue Decelles,
Montréal, Québec, CANADA H3T 1V6

**Centre de recherche sur les transports, Université de Montréal, C.P. 6128, succursale
A, Montréal, Québec, CANADA H3C 3J7

We describe a new heuristics, based on the Tabu Search approach introduced by Glover, for solving Vehicle Routing Problems. This heuristic is extremely flexible and can handle with minimal adaptation several types of constraints on the vehicle routes: vehicle capacity, maximum length of routes, etc.

The central element of the proposed algorithm is GENIUS, a generalized insertion heuristic for Traveling Salesman Problems developed by the authors.

The heuristic has been tested extensively on classical instances of the VRP, such as Christofides and Eilon's problems. Computational results show that this new method compares extremely favorably with the best existing heuristics for the VRP.

Session 18
Logistics – Logistique

Président / Chairperson:	Jacques Roy
Date:	Monday, June 10 Lundi, 10 juin
Heure / Time:	11:00
Lieu / Location:	Château Frontenac Salon Frontenac

The Location-Allocation-Inventory-Routing Model in the Design of Strategic Distribution Systems

Marc Goetschalckx, Sunghun Song and Mike Cole

School of Industrial and Systems Engineering, Georgia Institute of Technology,
765 Ferst Drive, Atlanta, GA 30332–0205, U.S.A.

Increasing competitive pressures and accelerating geo-political changes require an ever increasing attention to the rapid design and evaluation of global industrial logistics systems.

Accurate models for such strategic distribution systems must incorporate all contributing costs. The LAIR model will be presented, which is the first one to incorporate Location, Allocation, Inventory, and vehicle Routing costs in strategic distribution systems. A heuristic solution method for the LAIR model will also be introduced. The solution identifies, in addition to the traditional location and allocation decisions, the proper safety stock level and location, the correct transportation channel and transportation quantity, and the most economical local delivery method.

The LAIR model is one of the solution algorithms in the Computer Integrated Modeling and Planning Environment for Logistics or CIMPEL. The CIMPEL environment is oriented towards rapid prototyping, graphical display, and interactive sensitivity analysis. Other modules allow the design of fixed routes and linehaul-backhaul routes. The design philosophy, overall program structure, and implementation details of CIMPEL will also be discussed.

Continuous Modeling in Logistics

André Langevin

Département de mathématiques appliquées, École Polytechnique, Université de Montréal,
C.P. 6079, succursale A, Montréal, Canada H3C 3A4.

PART I. A survey of continuous models in logistics

Continuous models are being used: 1) to model integrated logistic systems, 2) to devise robust models necessitating few data, and 3) to approximate discrete models. This paper reviews more than sixty articles published mainly in the last two decades on continuous modeling in the field of logistics. A taxonomy of problems is proposed. Problems are classified according to decision variables, type of objective functions and system parameters (topological structure, metric, numbers of vehicles, of depots, type of servicing strategy, etc.).

PART II. The use of continuous models in multiple-vehicle delivery zone design

This paper addresses the problem of effective use of continuous models in planning an urban delivery service. Newell and Daganzo (1986) and Langevin and Soumis (1989) have proposed methods for the design of delivery zones based on analytic formulas. The objective was to determine optimal shape and dimensions of the zones. In both papers, one can find an analysis for a single zone at a fixed distance from the depot and then a procedure to partition a circular region with the depot at the center into zones, each one assigned to a single vehicle. Although useful for theoretical studies, the proposed procedures could hardly be used in practice: no city of course is circular, the depots can be located anywhere and, geographical constraints as well as a specific network of highways have to be taken into account.

This paper presents a prototype of an interactive decision support system that integrates the analytic formulas derived by the preceding authors. The main feature of the system is to enable the planner to interactively construct each zone on the screen while being guided by a displayed optimal shape which is function of the distance from the depot, the vehicle speed and capacity, the duration of the route, etc. The planner can thus incorporate the knowledge he has of the region, which is difficult to model otherwise, into the designing process. Hence one can take into account all the geographical peculiarities. The system uses functions of spatial distribution of the customers over the region rather than their exact locations.

An experimental study on zone planning for a company offering a courier service in the Montreal metropolitan area has been conducted and the results are presented in this paper. Relevancy of solutions and practical consideration issues are discussed.

The use of continuous models can be useful when planning a new service or a possible expansion of an existing one. As they rely more on distribution functions than exact locations of the customers, they do not need as much data as integer programming models. Their development and implementation are therefore easier and faster to achieve. They could be particularly useful in developing countries where resources in hardware and expertise are limited.

The main contribution of this paper is to show how continuous models can be efficiently used in practice. Even though they are only approximations, they bear sufficient information to be very useful at the strategic and tactical levels of planning.

References

Langevin and Soumis (1989) "Design of Multiple-Vehicle Delivery Tours Satisfying Time Constraints", *Transportation Research.*

Newell and Daganzo (1986) "Design of Multiple-Vehicle Delivery Tours: A Ring-Radial Network", *Transportation Research.*

Planning Models for the Distribution and Transportation of Containers

Teodor Gabriel Crainic[1,2], Michel Gendreau[2], and Pierre Dejax[3]

[1] Département des Sciences administratives, Université du Québec à Montréal, C.P. 6192, succursale A, Montréal, Canada H3C 4R2
[2] Centre de recherche sur les transports, Université de Montréal, C.P. 6128, succursale A, Montréal, Canada H3C 3J7
[3] Laboratoire Économique, Industriel et Social, École Centrale de Paris, 92295 Châtenay-Malabry Cedex, France

---

The planning and management of the land transportation part of the international maritime shipping of containers is an extremely complex activity, especially if one aims to simultaneously optimize the cost and the service aspects of operations in a competitive environment. Furthermore, the high cost of acquiring, maintaining, handling and transporting containers make the issue of adequate container management an extermely relevant problem. In this context, Operations Research models and methodologies offer very useful tools that a company may use to analyze, evaluate and plan its strategic and operational policies.

The objective of this paper is to review the main issues related to the management of a multimodal multicommodity container land transportation and distribution system, and to present an overview of the models and methodology that we propose for its global, integrated planning. The presentation is articulated around four main parts.

We first present the general characteristics of a container land transportation and distribution system, discuss the main planning issues, and present the integrated, multilevel methodological framework that we propose for their modeling and analysis.

The second part is dedicated to a brief review of our latest results in solving the model aimed at the *strategic/tactical* planning of operations. The main issues that are considered at this level are the choice of the depots, the allocation of customers to depots for each type of container and for each direction of movement, and the determination of the major interdepot flows of empty containers. The planning horizon of these decisions generally covers a six to twelve-month period. The main data requirements are the global demand forecasts, plus the general characteristics of the system, possible depot sites, fleet size and composition, etc., suitably aggregated. The so-called (strategic/tactical) *transportation plan* is the main output of this stage. The transportation plan will guide the actual operations of the system. It is also a privileged way of evaluating, in terms of system performance and network and fleet utilization, the consequences of various strategic scenarios concerned with the evolution of the company and with variations in its environment: demand, competition, laws and regulations, etc.

The strategic/tactical model may be formulated as a multimode, multicommodity location/distribution problem with interdepot balancing requirements. We have completed, during the last twelve months or so, the study of several methological approaches for solving this model: a dual ascent heuristic, several Lagrangean based methods, various greedy heuristics, etc. We report the main conclusions of these studies, as well as comparisons with results obtained by using branch-and-bound (sequential and parallel) and tabu search algorithms.

The third and fourth parts are dedicated to the day-to-day, operational planningof the company's activities. At this level, one has to make sure that demand is satisfied and that the most effective routes and means of transportation are selected and used. Major factors that characterize this level of planning are the dynamic and stochastic environment and the guidelines established by the strategic/tactical plan. The main question which is relevant for this problem is:

- How to achieve the management and dispatching of the fleet of containers, loadedand empty, in order to satisfy the known and forecasted demand and to perform the planned interdepot empty container movements, while considering the uncertainties associated with supply, demand and transportation, and the fact that transportation costs and reliability may vary considerably due to the use of many independent local carriers?

Other important questions, associated to the previous one, are:

- Which level of inventory of empty containers of each type should be maintained at each depot to face unexpected demands?
- How to take advantage of possible substitutions between container types to satisfy the demand?
- How to choose the transportation mode and carriers to undertake the shipment of loaded and empty containers between ports, or depots, and customers and vice versa, or between two depots?
- How to jointly manage the distribution of loaded and empty containers and the routing of the vehicles which carry them?

To address this problem, we formulate an *empty allocation problem* and an *empty and loaded routing problem*, and develop a separate optimization model for each of them. These two models are then linked together into a comprehensive decision support system for the operational management of the fleet, which is, in turn, connected to the strategic/tactical model and the company's M.I.S.

The allocation model aims at the determination of the "best" distribution of empty containers that satisfies both known and forecasted customer demands. Economic criteria guide the search for solutions which have to respect the strategic/tactical objectives and policies of the company. To allow for the high level of uncertainty inherent in the real-time operation of a transportation system and for the impact of current decisions in the future, it is a stochastic, dynamic network model which covers a given planning horizon, and which is normally used as a rolling horizon model of which only the first period (say one day) decisions are implemented. We briefly present the modeling framework, compare it to models developed for similar problems for different transportation modes, and discuss various algorithmic and data requirements issues.

The routing model strives to minimize the overall transportation cost of the loaded and empty containers from their origin to their destination. The inputs to this model are the list of empty movement requests produced by the allocationmodel and a list of loaded movement requests coming from the company's M.I.S., as well as detailed data on the available transportation alternatives. These data include cost, distance, and time information for the various possible transportation modes (rail, road, mixed, etc.), the different carrier contract options (one-way, two-way, circuit, etc.), the transportation regime for each movement (which indicates whether the customer or the shipping company is responsible for it), and the specification of customer transportation arrangements with regards to mode (direct link with the railway network, for example) and to pick-up and delivery schedules. Particular requirements concerning the associations of container types and carrying vehicle characteristics have also to be indicated. It is worthy to note that, since the container shipping company does not own, in any significant measure, the required land transportation means, substantial savings can be achieved by using a routing module. This is particularly true fortruck transportation where most savings are achieved by generating multi-stop "circuit" truck routes and by suitably matching requested container movements to these routes. We describe the general modeling and algorithmic framework that we propose, based on set covering and column generation principles, and discuss the various specializations of this framework according to the time horizon which is contemplated and to the type and number of transportation companies which have to be simultaneously considered.

We conclude by presenting an overview of the proposed decision support system, and by identifying future research directions.

Session 19
Focus 2

Président / Chairperson:	Stavros A. Zenios
Date:	Monday, June 10 Lundi, 10 juin
Heure / Time:	14:00
Lieu / Location:	Château Frontenac Salon Frontenac

Interior Point Methods for Linear Programming: Computational Results

Irv Lustig

Department of Civil Engineering and Operations Research, Princeton University, Princeton, NJ 08544, U.S.A.

Since the introduction of Karmarkar's algorithm in 1984, a flurry of research activity has occurred making improvements to the original interior point method. This activity has been primarily theoretical. However, good implementations of some interior point methods have been produced, and very large-scale linear programs with tens of thousands of rows and tens of thousands of columns have successfully been solved on scientific workstations, thus extending the range of solvable linear programs. In this talk, a short history of the development of practical interior point methods for linear programming will be given. This will include a background for understanding a primal-dual predictor-corrector method. Recent computational experience will be presented for solving large-scale linear programs with a mature implementation of this latter method. Comparisons to a state-of-the-art implementation of the simplex method will be given, indicating the promise for using interior point methods for solving very large problems. In addition, experience in solving a 12.75 million variable linear program that arises from an airline crew scheduling application will be presented.

Session 20
Air Transportation – Transport aérien

Président / Chairperson:	Mark S. Daskin
Date:	Monday, June 10 Lundi, 10 juin
Heure / Time:	15:15
Lieu / Location:	Château Frontenac Salon Frontenac

Dynamic Ground-Holding Strategies for Air Traffic Control

François Soumis* and Amedeo R. Odoni**

* GERAD and Ecole Polytechnique, C.P. 6079, succursale A, Montréal, CANADA H3C 3A4
**Operations Research Center, Room 33–404, Massachusetts Institute of Technology,
Cambridge, MA 02139–4307, U.S.A.

Ground-holding (or "gate-holding") of aircraft is becoming increasingly common throughout the world as air traffic congestion grows. Ground-holding is the practice of delaying the departure of a flight because of anticipated congestion at its airport of destination. The motivation for doing so comes from the fact that it is both less expensive and safer for an aircraft to wait on the ground, prior to take-off, than in the air.

The problem of developing strategies for which aircraft to hold on the ground and for how long (the "ground-holding policy problem", GHPP) is a difficult one. The GHPP is: combinatorial, because there is a large number of landing periods ("slots"); stochastic, because there is often a large amount of uncertainty about the acceptance rates ("landing capacity") of the airports of destination; and dynamic, because the state of information regarding airport capacity and expected delays is updated throughout a day and decisions on whether to hold or release a particular departure are made individually at the time when each departure is scheduled. Because the extensive use of ground-holding is only a relatively recent phenomenon, the strategies currently used in practice by the FAA and other ATC organizations are rather simplistic and there are strong indications from computational experiments that more sophisticated solutions to the GHPP can result in very large cost savings to the airlines and their passengers.

In this paper, we shall present the results of research currently in progress aimed at solving large-scale instances of the GHPP. Two non-linear network models that take into account the principal stochastic aspects of the problem will be described. Time in the models is discretized and for each time interval the distribution of the forecasted capacity is explicitly considered.

The nodes of the network represent (i) incoming flights and (ii) landing time intervals. The arcs correspond to the assignment of flights to time intervals, with and without delay. The cost function includes a linear term for ground-holding delay (before take-off) and a non-linear term for the expected airborne waiting at the end of each time interval. The number of waiting airplanes at the end of each interval is computed as a function of: the number of airplanes arriving during the interval; and the stochastic landing capacity during the interval. The underlying assumption is that the number of aircraft in queue is small relative to the number of new arrivals during an interval of time.

The first of our models approximates the random number of aircraft waiting to land by its mean value. The second model approximates this random number by a random variable whose distribution is known.

We shall discuss solution approaches to these models and also anticipate having some computational examples.

Airline Seat Allocation with Multiple Nested Fare Classes

S.L. Brumelle* and J.I. McGill**

* Faculty of Commerce and Business Administration, University of British Columbia,
Vancouver, CANADA V6T 1W5

**College of Business, University of Denver, Denver, CO 80208, U.S.A.

This paper addresses the problem of determining optimal booking policies for multiple fare classes sharing the same seating pool on one leg of an airline flight when seats are booked in a nested fashion and when lower fare classes book before higher. 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 appropriate for either the discrete or continuous demand cases. These conditions are further simplified to a set of conditions relating the probability distributions of demand for the various fare classes to their respective fares. The latter conditions are guaranteed to have a solution when the joint probability distribution of demand is continuous. 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 approximation solutions obtained by P. Belobaba using the expected marginal seat revenue (EMSR) method.

Session 21
Operations – Exploitation

Président / Chairperson:	Amedeo R. Odoni
Date:	Monday, June 10 Lundi, 10 juin
Heure / Time:	16:30
Lieu / Location:	Château Frontenac Salon Frontenac

A Hierarchical Solution for the Service Network Design Problem

Judith M. Farvolden

Department of Industrial Engineering, University of Toronto, Toronto, Ontario, Canada H5S 1A4

The motivating application for this paper is a Service Network Design Problem (SNDP) which arises in the Less-Than-Truckload motor carrier industry. Reduction of inventory levels and satisfaction of customer demand in a timely fashion necessitates the coordination of the production and transportation functions in a logistics system. To satisfy the increasingly sophisticated demands of shippers, the carriers must incorporate detailed scheduling into their load planning operations.

The distribution system can be modeled by a scheduling network, $G = \{N, L\}$, where the nodes, N, represent the terminals at each point in time over a planning horizon and the links, $(i,j) \in L$, represent inventory and transportation movements. The scheduled set of services determines the number of vehicle departures, y, between each pair of terminals in the distribution system, at each point in time. The decision to dispatch a vehicle creates a capacitated link in the network. Taken together, the scheduled set of services results in a schedule map, over which the known set of shipments, S, must be distributed, and creates net supplies of and demands for vehicles, d, at each terminal. Vehicles travel empty between terminals, to rectify any flow imbalances.

The SNDP is modeled as a mixed integer linear program which captures the system-wide interactions between the vehicle scheduling and shipment routing decisions:

$$\text{minimize } F(x,y,z): \qquad \sum_{s\in S}\sum_{ij\in L} c^s_{ij}\, x^s_{ij} + \sum_{ij\in L} t_{ij}\, y_{ij} + \sum_{ij\in L} q_{ij}\, z_{ij}$$

$$\text{subject to}: \qquad \sum_{j\in N} x^s_{ji} - \sum_{j\in N} x^s_{ij} = f^s_i \qquad \forall i \in N,\ \forall s \in S$$

$$\sum_{s\in S} x^s_{ij} \le y_{ij}\, W \qquad \forall (i,j) \in L$$

$$\left(\sum_{j\in N} y_{ij} - \sum_{j\in N} y_{ji}\right) - \left(\sum_{j\in N} z_{ij} - \sum_{j\in N} z_{ji}\right) = d_i \qquad \forall i \in N$$

$$x^s_{ij} \ge 0 \qquad \forall (i,j) \in L,\ s \in S$$

$$y_{ij} = 0, 1, 2 \ldots \qquad \forall (i,j) \in L$$

$$z_{ij} = 0, 1, 2, \ldots \qquad \forall (i,j) \in L$$

where:

$$f^S_i = \begin{cases} -w^s & \text{if node } i \text{ is the origin node } q^s \\ -w^s & \text{if node } i \text{ is the destination node } r^s \\ 0 & \text{otherwise} \end{cases}$$

The objective is to determine the distribution pattern, x, for all shipments and the schedule of loaded movements, y, and empty movements, z, which minimize total cost. The decision variable x^s_{ij} is the flow on the s^{th} shipment on link $(i, j) \in L$ with associated link cost c^s_{ij}. The cost of transportation is a function of the cost of the service, t_{ij}. The link cost c^s_{ij} thus represents the level of service penalty associated with sending the shipment via this service. The decision variable z_{ij} is the integer number of empty vehicles traveling on link $(i, j) \in L$, at cost q_{ij}.

The flow is subject to three sets of constraints. The first ensures that the shipment, of size w^s will be transported over the service network from its terminal of origin q^s to its destination terminal, r^s. The second set limits the sum of the flows on each link to the capacity $y_{ij}\ W$, which is a function of the number of scheduled departures and the size of the vehicles, W, in the homogenous fleet. There is also a balance of flow of scheduled empty and loaded vehicle departures, d_t at each node. The dual variables associated with the constraints of the above program are μ^s_i for all $i \in N$ and $s \in S$, λ_{ij} for all $(i, j) \in L$, and γ_i for all $i \in N$ giving the marginal values of an additional unit of a shipment, of link capacity and a vehicle respectively.

Considered simultaneously, all of these decisions result in a problem too large to be solved in a reasonable amount of time. Examination of the constraint matrix of SNDP indicates that for a given set of vehicle departures, y, the problem decomposes into a shipment distribution problem in x and an empty vehicle balancing problem in z. The linear program in x has the form of a multicommodity network flow problem (MCNF).

The MCNF can be recast in terms of path flows and solved by primal partitioning and decomposition. The partitioning is performed on an arc-chain incidence matrix of MCNF, similar within a change of variables to the constraint matrix of the master problem generated in a Dantzig-Wolfe decomposition, to isolate a very sparse, near-triangular working basis of greatly reduced dimension. The variables in the working basis are paths for shipments split over multiple basic paths. The rows are the flow conservation constraints for those split shipments and the capacity constraints for saturated links. Only the operations on the sparse, near-triangular working basis require matrix algebra. The majority of the simplex operations performed on the partitioned basis are simply additive and network operations. The algorithms solves MCNF with the properties of the shipment distribution problem on average 20 times faster than an implementation of Dantzig-Wolfe decomposition and up to 100 times faster than a straightforward linear programming solution.

The remaining scheduling problem in y is an integer program which falls into the class of NP-hard problems for which exact solution techniques are prohibitively expensive. In order to account for the system impacts of scheduling decisions, the heuristics must guide the solution along descent directions determined by the algebra of the SNDP. Optimal duals of the MCNF associated with link capacity constraints provide a measure of the change in the objective function if the allocation of vehicles to that link is changed. From these, subgradients which define directional derivatives along which the objective function decreases have been derived.

The introduction of a new service is modeled by increasing the capacity of a link from zero to one and the cancellation of a service by decreasing the capacity of a link to zero. In either case, a new network and a new set of candidate nonbasic paths with modified path cost, p_{sk}, are created. (The modified path cost is the sum of modified link costs for all links in the path. A link cost is modified by the nonzero dual variable associated with a saturated link.) Flow will be diverted to a new service if the reduced costs of at least one new path is negative. The subgradient associated with a service introduced on link $(i,j) \in L$ is

$$c_{ij} = \arg\min_{s \in S} \{\overline{p}_{sk'}\}$$

where the reduced cost of a nonbasic path is $\overline{p}_{sk'} = \tilde{p}_{sk'} - \tilde{p}_{sk}$ for unsplit shipments currently assigned to a path with modified cost $\tilde{p}_{sk}$, and $\overline{p}_{sk'} = \tilde{p}_{sk'} - \mu_i^s$ for split shipments.

In the case of a cancelled service modeled by an unsaturated link, $(i,j) \in L$, the subgradient is again simply the minimum reduced cost over all candidate paths calculated according to $\overline{p}_{sk'} = \tilde{p}_{sk'} - \tilde{p}_{sk}$ that pulls flow off link (i,j) onto the path that does not include (i,j), if none of the shipments on the link are split. The calculations involving split shipments are more complex. It is well known that the dual variable associated with a saturated link is a measure of the change in the objective function if that capacity is reduced.

Preliminary numerical results indicate that the subgradients often provide reasonable estimates of the change in the objective function, but that in a few cases network interactions mitigate against a redistribution of flow with lower cost. Rules to generate candidate paths with minimal interaction are proposed since interactions occur between saturated links and the solution is most certainly highly degenerate. The empty balancing subproblem is readily solved as a standard minimum cost network flow problem.

While the shipment routing subproblem can be solved to provide optimal dual variables to be used in the service scheduling master problem, the time to solve the MCNF is considerable, thus undermining the utility of a solution in an operational setting. However, the nature of the optimal solution to the MCNF is well understood. The majority of shipments will follow a single basic path, and only a few will be split over multiple basic paths if a service, represented by a link in the network, becomes saturated. A near optimal solution provides an estimate of the optimal dual variables which are used to improve the vehicle scheduling solutions by multiplier adjustment methods.

The SNDP has been partitioned into a set of problems with well-known structure, about which there is considerable knowledge of solution procedures and the nature of the solutions. The decisions made in each problem interact and cannot be made in isolation. Vehicle scheduling decisions determine the constraints in the shipment routing and empty balancing subproblems, and the shipment routing problem provides information about the cost of the current scheduled services. With a linking mechanism to transfer information between the problems, the SNDP can be solved by a hierarchical solution method which iterates between the vehicle scheduling master problem and the shipment routing and empty balancing subproblems.

An Analytical Model for Optimal Design of an Automated Guided Vehicle System

M. Eric Johnson and Margaret L. Brandeau

Department of Industrial Engineering and Engineering Management, Stanford University, Stanford, CA 94305, U.S.A.

In their brief 30–year history, fixed path material transport systems, such as Automated Guided Vehicle Systems (AGVSs), have progressed from driverless tractor systems of the 1950's to sophisticated computer controlled vehicle systems with powerful on-board computers. However, while vehicle guidance technology has seen remarkable progress in the last 20 years, little is yet known about the design and operation of efficient AGVS's. Almost all research to date on the design of such material transport systems has focused on simulation studies.

Many transportation research problems that have been extensively studied are related to AGVS problems. These include vehicle routing and scheduling, dial-a-ride problems, and network design and transportation planning models. However, the missing element in all of these problems is both vehicle and queueing congestion. In multi-vehicle systems, the possibility of vehicle congestion along routes exists. Additionally, the stochastic nature of most manufacturing systems gives rise to fluctuations in demand for transportation, causing queueing congestion of material waiting to be transported. The extent to which such congestion affects the performance of the system varies greatly with the complexity, size, design, and operation of the network.

In this paper we develop and solve an analytical model for the optimal design of an AGVS that explicitly incorporates congestion effects. Our model, based on an actual AGVS design problem faced at Hewlett-Packard, is also applicable to other asynchronous fixed path material handling systems. The problem we consider is that of adding an AGVS to an existing material handling system. The objective is to determine which stations to include in the network to maximize the benefit of the network, subject to a constraint that the average waiting time in the system not exceed a predefined limit. In the absence of an analytical model, one would have to simulate all possible combinations of stations and routes in order to find the configuration that maximizes the benefit of the AGVS. Our contribution is to solve the problem using an analytical model. We present a model and solution for a single vehicle system, and then describe our ongoing research on multi-vehicle systems.

Single-Vehicle System. Based on an actual problem faced at Hewlett-Packard, we formulate the single vehicle AGVS design problem as a binary integer program with a linear objective, a set of linear constraints, and one nonlinear constraint. The (binary) decision variables $(y = \{y_{ij}\})$ indicate whether demand for material transport from node i and node j will be handled by the AGVS; if not, it is assumed that the demand can be handled by an existing, non-automated system, at a higher cost. The objective is to determine a set of routes (and associated pickup/dropoff stations) to include in the network to maximize the dollar benefit of the automated transport (calculated as direct labor savings) minus the cost of pickup/dropoff stations and the cost of the vehicle and control equipment. A set of linear constraints ensures that if material is transported from node i to node j, the cost of pickup/dropoff stations at node i and j is charged. The model also incorporates one nonlinear constraint which ensures that the expected time an order must wait for transport (the expected time the vehicle is busy transporting other orders $(E(Q(y)))$ plus the time the vehicle travels empty from its last destination to the current order $(E(T_E(y)))$ does not exceed a prespecified limit (K). Since the cost of the guide path in an AGVS is typically negligible in comparison with the cost of pickup/dropoff stations and the cost of the vehicle, we assume that any needed guide paths will be available at no cost; thus, we do not need to consider the route connectedness constraints.

We develop an analytical expression for the expected time an order must wait for transport $(E(Q(y)) + E(T_E(y)))$, as a function of the y_{ij}'s, in the following manner: We assume that orders are handled on a first-come-first-served (FCFS) basis. Then (as a function of the y_{ij}'s) we calculate the full trip distribution (the probability that a random order is from node i to node j) and the empty trip distribution (the probability that the last order was delivered to node i and the next order originates from node j). We use these probabilities to calculate expected empty travel time of the vehicle to pick up an order $(E(T_E(y)))$, and expected service time for an order (time spent travelling empty to pick up the order, load time, time spent transporting the order, and unload time). To calculate expected time until the vehicle is free $(E(Q(y)))$, we make the additional assumption of independence of successive service times. This assumption is violated in the AGV network because the probability distribution of the current service time depends on the previous service time: for example, knowledge of a particularly long service time may signal the completion of a delivery to a remote node of the network, thus changing the probability distribution of the next service time. Given this simplifying assumption, we develop an exact expression for $E(Q(y))$ for the case of Poisson order arrivals (i.e., an M/G/1 system); and for the case of a general distribution of order arrivals (i.e., a G/G/1 system) we use an approximate formula for mean queue delay to develop an expression for $E(Q(y))$. Our computational experience indicates that when relative distances between nodes are within the same order of magnitude, expected queue congestion calculated using our queueing formulas is within one or two percent of that calculated using simulation.

In order to solve the optimal design problem (a linear BIP with a set of linear constraints and one nonlinear constraint), we develop a relaxed integer version of the problem that replaces the nonlinear constraint with a looser linear mean waiting time constraint (the new linear constraint may allow solutions that violate the original nonlinear constraint). The linear constraint is a reasonably tight approximation of the original nonlinear constraint, and relies partly on our establishment of the following important bound: the expected amount of time the vehicle speeds travelling empty (independent of relative distances between nodes, relative demands at different nodes, and the set of orders selected for transport) must always be at least half the expected amount of time the vehicle spends travelling full. We solve the resulting linear problem with branch and bound, using a further LP relaxation to bound. To check feasibility of any integer solution we obtain, we examine the original nonlinear constraint.

We then use the model to solve the original application faced at Hewlett-Packard. We solve the problem for different values of K (the upper limit on mean waiting time for transport) in order to examine the cost-benefit tradeoff of different solutions, and then identify the best solution. This solution, approved by planners at Hewlett-Packard, has not yet been implemented because the planned AGVS has not yet been installed; however, we hope that by the time of the TRISTAN conference, the system will have been installed.

Two key simplifying assumptions of our model are independence of successive service times (used to obtain an analytical formula for vehicle congestion) and FCFS order priority. For the case of a general distribution of order arrivals, another simplification is the use of an approximate analytical formula for

$E(Q(y))$. We show that the model can be extended in the following manner to eliminate the restrictions caused by these assumptions: We solve the model for different values of K, starting with values of K larger than we may actually be willing to accept. When verifying feasibility of an integral solution obtained in the branch and bound process, instead of substituting the solution into the original nonlinear constraint, we perform simulation to determine whether or not the solution is actually feasible (without the restrictive assumptions of the analytical model). In this manner, the analytical model is used as a guide to identify potentially optimal solutions, so that simulation must be performed only for a very limited set of potential solutions. We prove that this use of the model will yield the benefit-maximizing solution for the actual problem (i.e., the analytical model will not overlook a potentially optimal solution). We also present computational experience.

Multi-Vehicle System. Finally, we propose to present the results of our ongoing research on optimal design of multiple-vehicle AGVS's. Our research has thus far focused on "zoned" networks in which individual vehicles are limited to certain areas of the network, and zones of different vehicles have only a limited amount of overlap. The question is to determine no only which routes and stations to include in the AGVS, but also vehicle zones in the network, in order to maximize benefit, subject to a waiting time constraint. In this case, our formula for expected waiting time incorporates the additional factor of potential vehicle congestion along arcs of the network that are shared by more than one vehicle. We expect to have made significant progress on the multiple-vehicle model by the time of the TRISTAN conference.

Using Game-Theoretic Methods in Decision Support Systems for Traffic Scheduling

A.S. Belenky* and M.S. Dubson**

* Institute of Control Sciences, U.S.S.R. Academy of Sciences, Moscow, U.S.S.R.
**Central Institute of Mathematics and Economics, U.S.S.R. Academy of Sciences, Moscow, U.S.S.R.

Among different optimization-based approaches to problems of traffic scheduling, the one based on game-theoretic methods looks rather uncommon. Except for several papers attempting to formalize the decision making process in transportation systems in terms of matrix games, no examples successful practical application of game approach to planning of transportation can be found in literature, at least to the authors' knowledge. The circumstances which, to an extent, explain the paltry amount of work in this area, are the difficulty of interpreting the solution of a game in terms common for a practicing transport worker, the tradition to view the game-theoretic approach primarily as a means for a qualitative analysis of systems (transport systems in particular), the general problems with developing software intended for transport end users, and also the existing tendency for a decline in the popularity of the optimization methods which is obviously connected with a fast progress of expert systems and the widespread use of personal computers. However, it is evident that the potentialities of such a powerful tool as the game methods, are far from having been exhausted and new results in this area, especially those directed toward practice, may be of considerable interest.

In [1], the problem of tariffing the seaborne freight traffic is considered as a zero-sum two-player game on polyhedral sets M and W with the payoff function

$$\max_{u \in H} < u, y > + < y, Dx > + < p, x >$$

where M is a convex polyhedron in R^m, W, $H \subset R^n$ and D is a matrix with real elements. Necessary and sufficient conditions for solvability of such games are obtained and it is shown that the components of a saddle point $(y^*, x^*) \in W \times M$ can be found by solving quadratic and linear programming problems whose systems of constraints are specified by some linear-algebraic transformations of the matrices and vectors involved in the definition of the sets M, H, and W. The components of saddle points admit a very natural interpretation in terms of transportation tariffs and volumes.

Practical implementation of the method within the framework of a decision-making support system calls for developing special end-user-oriented software for the available PCs.

In [2], a specialized interactive system has been presented which is intended for solving planning problems based on linearly-structured mathematical models and is equipped with a considerable number of service functions. The system has been successfully tested on numerous practical problems of planning production processes and has won a high appraisal of end users.

In this paper, we propose a system which combines the idea of using the game-theoretic approach to traffic scheduling with the proven ideas exploited in the design of flexible end-user-oriented software for

PCs. The system suggests much service for the end user who is not supposed to be an expert in computers or in mathematics of optimization.

The basic data for the model (elements of the matrices and vectors specifying the polyhedral sets M, H, and W) can be set up by using any databases, in particular, the ones designed *ad hoc* by the user. The system provides for spreadsheet imports in the formats of popular applied packages such as Lotus, dBase, Rbase, Framework, Quattro, Supercalc, etc. A special version of the spreadsheet processor with vector and matrix operations is available. An important point is that, while setting up the basic information, the user can analyze some of the data graphically, which enables him (or his experts) to estimate the dynamics of the parameters and to forecast, if necessary, their future values. To this end, the user can, in particular, employ modern packages for the analysis of time series, such as Mesosaur, or Trend, or SPSS.

In accordance with the method used to solve the game under consideration, the system generates auxiliary pairs of quadratic and linear programming problems; the systems of constraints of these problems are generated as mentioned above, using GAMS package to perform linear-algebraic transformations of matrices and vectors. Also, the optimization modules of GAMS are used for solving the auxiliary problems. Due to the fact [1] that solvability of the game is equivalent to that of the set of auxiliary problems, the user can check whether the input data are feasible (in the sense of the solvability of the corresponding game) and adjust the data so as to make the game feasible. Thus, the user actually gets a means for analyzing the game and solving it in the course of the interactive process.

To represent the computational results in an easy-to-use form, they can be processed by some report generators, for example, by those provided by GAMS, or else by the interface modules for the data export in the formats of the above mentioned popular spreadsheets packages. It is well known that these packages provide the user with high-quality interactive business graphics which substantially facilitates the decision-making process.

References

[1] A.S. Belenky (1986) "An Antagonistic Game on Polyhedral Sets", *Automation and Remote Control* 47(6), 757–761.

[2] E.P. Borlsova, M.S. Dubson, and E.G. Golshtein (1990) "An Interactive System for the Analysis of Multicriteria Problems", *Ekonomika and Matem. Metody*, 26(4), (in Russian, to be translated in *Matecon*).

Session 22
Probabilistic Analysis for VRP Algorithms – Analyse probabiliste des algorithmes pour le VRP

Président / Chairperson:	Harilaos Psaraftis
Date:	Tuesday, June 11 Mardi, 11 juin
Heure / Time:	09:00
Lieu / Location:	Château Frontenac Salon Frontenac

Probabilistic Analysis of Algorithms for the Capacitated Vehicle Routing Problem with Unsplit Demands

David Simchi-Levi and Julien Bramel

Department of Industrial Engineering and Operations Research, Columbia University,
New York, NY 10027, U.S.A.

The Capacitated Vehicle Routing Problem can be stated as follows: A set of customers has to be served by a fleet of identical vehicles of limited capacity q. The vehicles are initially located at a given depot. Associated with each customer is a nonnegative demand, which is the amount of load that must be delivered to that customer. The capacity constraint states that the total amount of load delivered by any one vehicle cannot exceed a prespecified number q. The objective is to find a set of routes for the vehicles, where each route begins at the depot, visits a subset of the customers, and returns to the depot, such that the vehicle capacity is not exceeded and the total length of the routes is as small as possible.

The capacitated vehicle routing problem belongs to the class of NP-hard problems which indicates that the existence of a polynomial time algorithm to solve the problem optimally is unlikely. Consequently, in the last five years, a great deal of work has been devoted to the development and theoretical analysis of heuristic algorithms for different versions of the problem. Specifically, what we refer to as theoretical analysis includes worst-case performance analysis and probabilistic analysis. While the former establishes the maximum deviation from optimality that can occur for a given heuristic algorithm, the latter establishes the average performance of the heuristic under assumptions on the distribution of the data.

Altinkemer and Gavish (1985) and Li and Simchi-Levi (1990) perform worst-case analysis for different versions of the problem. Haimovich and Rinnooy Kan (1985) perform probabilistic analysis of a class of heuristics for the special version of the problem in which the demand of a customer can be split over more than one vehicle. An excellent survey of these works appears in Haimovich, Rinnooy Kan and Stougie (1988).

In this paper we perform probabilistic analysis of the capacitated vehicle routing problem with the additional constraint that a customer's demand may not be divided over more than one vehicle. The problem is called the capacitated vehicle routing problem with unsplit demands. As Haimovich and Rinnooy Kan aptly put it, in this case *"... the problem is much harder; this additional constraint introduces bin packing features into the routing problem and an appropriate solution method is not easily found."*

Our analysis will use results from the bin packing problem, a problem that has been analyzed extensively in the literature. An instance of the bin packing problem is composed of the bin capacity (equal to 1) and a set of items each with a prespecified size no larger than 1. The problem is to find the smallest number of bins in which these items can be packed, subject to the constraint that the total size of items assigned to a bin cannot exceed 1. For a recent survey on probabilistic analysis of heuristics for the bin packing problem, the reader is referred to Coffman, Lueker and Rinnooy Kan (1988).

In the capacitated vehicle routing problem with unsplit demands we assume, without loss of generality, that the vehicles' capacity q equals 1, and the demand of each customer is no more than 1. (As Coffman, Lueker and Rinnooy Kan explain, q is essentially a scale factor, so we can make the normalization $q = 1$).

Thus, vehicles and demands in the capacitated vehicle routing problem correspond to bins and items sizes, respectively, in the bin packing problem.

The customers and the depot are presented as a set of points on the Euclidean plane. We denote by $N = \{y_1, y_2, ..., y_n\}$ the set of customers, by w_k the demand of customer y_k, and by d_k the distance between customer y_k and the depot. In addition, let Z^* be the optimal solution value, i.e., the total distance traveled in the optimal solution of the capacitated vehicle routing problem.

Consider the demands $w_1, w_2, ..., w_n$ to be independent and identically distributed like Φ, where Φ is a probability measure defined on [0,1]. In this paper we find the asymptotic optimal solution value for any distribution of the demands Φ. This is done by showing that an asymptotically optimal algorithm for the bin packing problem, with item sizes distributed like Φ, can be used to solve the capacitated vehicle routing problem with unsplit demands.

Given the demands $w_1, w_2, ..., w_n$, let b^* be the number of bins used in the optimal solution for the bin packing defined by item sizes $w_1, w_2, ..., w_n$, and bin capacity equal to 1. As demonstrated in Rhee and Talagrand (1987), it is well known that there exists a constraint $\gamma > 0$ such that $\lim_{n\to\infty} \frac{b^*}{n} = \gamma$ almost surely, for any distribution of the item sizes.

The result we will prove in this paper is embodied in the following theorem.

Theorem 1.1: (Main Theorem) Let y_k, $k = 1, 2, ..., n$ be a sequence of independent random variables having a distribution μ with compact support in $\Re^2$ and suppose

$$E(d) = \int_{\Re^2} d(y)\, d\mu(y) < \infty.$$

Let the demands w_k, $k = 1, 2, ..., n$ be a sequence of independent random variables having a distribution Φ with support on [0,1], and let $\lim_{n\to\infty} \frac{b^*}{n} = \gamma$ almost surely. Then, almost surely,

$$\lim_{n\to\infty} \frac{1}{n} Z^* = 2\gamma E(d).$$

Thus, Theorem 1.1 fully characterizes the asymptotic optimal solution value of the capacitated vehicle routing problem, for any distribution function Φ. To prove this result, we start by presenting a lower bound on the optimal objective function. This lower bound is completely different from the lower bound developed by Haimovich and Rinnooy Kan (1985) for the case when the demand can be split.

We then present a **polynomial time** heuristic algorithm for the capacitated vehicle routing problem with unsplit demands based on a simple region partitioning scheme. We show that the cost of the solution produced by the heuristic converges to our lower bound for any distribution Φ, thus proving the main theorem of the paper.

An important consequence of Theorem 1.1 concerns the capacitated vehicle routing problem with unsplit demands when the distribution of the demands allows perfect packing, that is, when the wasted space in the bins tends to become a negligible fraction of the number of bins used. Formally, Φ is said to allow perfect packing if $\lim_{n\to\infty} \frac{b^*}{n} = E(w)$ $(a.s.)$, hence, for this class of distribution $\gamma = E(w)$. Karmarkar

(1982) proved that a non-increasing distribution (with some regularity) allows perfect packing. Rhee (1988) characterizes the class of distribution functions Φ which allow perfect packing. We show, surprisingly, that allowing the demands to be split or not does not change the asymptotic objective function value. That is, the capacitated vehicle routing problem with unsplit demands and the capacitated vehicle routing problem when demands can be split are **asymptotically equivalent**, when Φ allows perfect packing.

Finally, we report on computational experiments that we have conducted to verify the practical performance of our algorithm. These experiments show that the algorithm performs quite well for reasonable problem sizes.

Significance of the Contribution:

The contribution of our research to the field of vehicle routing is threefold. First, we obtain the asymptotic optimal solution value of the capacitated vehicle routing problem with unsplit demands for any distribution of the demands and when the customers are independently and identically distributed in a given region. This is especially important for the design of distribution systems, since it shows that the total cost behaves like the average radial cost to the depot. Second, we present a polynomial time heuristic algorithm that achieves this asymptotic optimal solution. Finally, by performing a series of computational experiments, we show that this algorithm also works well in practice, in addition to its nice theoretical properties.

References

[1] Altinkemer, K. and B. Gavish (1985) "Heuristics for Equal Weight Delivery Problems with Constant Error Guarantees", Working Paper Series No. QM8536, The University of Rochester.

[2] Coffman, Jr., G.S. Lueker and A.H.G. Rinnooy Kan (1988) "Asymptotic Methods in the Probabilistic Analysis of Sequencing and Packing Heuristics", *Management Science* 34, 266–290.

[3] Haimovich, M. and A.H.G. Rinnooy Kan (1985) "Bounds and Heuristics for Capacitated Routing Problems", *Mathematics of Operations Research* 10, 527–542.

[4] Haimovich, M., A.H.G. Rinnooy Kan and L. Stougie (1988) "Analysis of Heuristics for Vehicle Routing Problems", Golden, B.L. and A.A. Assad (eds.), *Vehicle Routing: Methods and Studies*, Elsevier Science Publishers, B.V., 47–61.

[5] Karmarkar, N. (1982) "Probabilistic Analysis of Some Bin-Packing Algorithms", Proceedings of the 23rd Annual Symposium, *Foundations of Computer Science*, 107–111.

[6] Li, C.L. and D. Simchi-Levi (1989) "Analysis of Heuristics for the Multi-Depot Capacitated Vehicle Routing Problems". To appear in *ORSA Journal on Computing*.

[7] Rhee, W.T. (1988) "Optimal Bin Packing with Items of Random Sizes", *Mathematics of Operations Research* 13, 140–151.

[8] Rhee, W.T. and M. Talagrand (1987) "Martingale Inequalities and NP-Complete Problems", *Mathematics of Operations Research* 12, 171–181.

On Refinements in Probabilistic Analysis of Geometric Problems

Patrick Jaillet

M.S.I.S. Department, CBA 5.202, University of Texas at Austin, Austin, TX 78712–1175, U.S.A.

In Beardwood et al. [1], the authors prove that for any bounded i.i.d. random variables $\{X_i : 1 \leq i < \infty\}$ with values in $\Re^2$, the length of the shortest tour through $\{X_1, ..., X_n\}$ is asymptotic to $\beta n^{1/2}$ with probability one (the same being true in expectation). This theoretical result has become widely recognized to be at the heart of the probabilistic evaluation of the performance of heuristic algorithms for vehicle routing problems. In fact it is used as the main argument in the probabilistic analysis of partitioning algorithms for the traveling salesman problem (TSP) by Karp [5]. Let us also mention the paper by Haimovich and Rinnooy Kan [2] in which a probabilistic analysis of a class of heuristics is performed for the capacitated vehicle routing problems.

Because of these algorithmic applications, results like that of Beardwood et al. have gained considerable practical interest. An important contribution on the subject is contained in Steele [7] in which the author uses the theory of independent subadditive processes to obtain strong limit laws for a class of problems in geometrical probability which exhibit nonlinear growth. Examples include the TSP, the Steiner, and rectilinear Steiner tree problem, and the minimum weight matching problem. Other problems of interest in transportation, such as the minimum spanning tree problem (MSTP), and some probabilistic versions of the TSP and MSTP, have been subsequently treated in different papers (see respectively Steele [8] and Jaillet [3]). For most of these problems, the results concern the almost sure convergence of the normalized means.

Questions about rates of convergence have been raised many times. There are in fact two issues concerning information on the rate of convergence: (i) What is the asymptotic size of $L_n - EL_n$? (ii) What can be said about the rate of convergence of the normalized means $EL_n/\sqrt{n}$ to β? For the TSP in the plane, Rhee and Talagrand [6] prove that, if the points are uniformly and independently distributed over the unit square, then there is a constant k such that $\|L_n - EL_n\|_p \leq k\sqrt{p}$ for each p. This means that L_n is very concentrated around its expectation, but it does not give any information on the value of EL_n. For the TSP in the plane, if one follows the usual subadditivity argument it is relatively easy to see that $EL_N/\sqrt{n} \geq \beta - c/\sqrt{n}$ for a positive constant c, where N is a Poisson random variable with parameter n (N is the number of points corresponding to a Poisson point process of intensity n times the Lebesgue measure over $[0,1]^2$). Also, Karp [5] has shown the following result:

$$EL_N/\sqrt{n} \leq \beta + 12/\sqrt{n}.$$

More recently, in [4], we have shown that such results could be given for the initial random process itself (and not only for its Poisson approximation). The main result for the TSP in the plane is then

$$\left|EL_N/\sqrt{n} - \beta\right| \leq 7/\sqrt{n} + 2/\sqrt{n-1}.$$

On the other hand, we do not know the best possible order of the "error term" (i.e., is $|EL_N/\sqrt{n} - \beta| = \Omega(1/\sqrt{n})$ or can we find a faster rate?). The same results and questions have been obtained for a large class of geometric problems (see [4]).

In order to made additional advances on these difficult questions, one has to conduct a much more detailed probabilistic analysis of the problems. Moreover, a level of generality such as in [4] seems difficult to maintain and progress seems to be possible only by considering the particularities of each specific problem.

In this talk, after a survey of the results obtained in [4], we will present our most recent findings on the TSP and MSTP with respect to rates of convergence and related topics. For example, among very recent results, we have been able to improve on (1) and show that for the TSP, $\beta \leq EL_N/\sqrt{n} \leq \beta + 6/\sqrt{n}$ (the lower bound was conjectured in Karp [5], but remained an open question). We have also characterized (in expectation) the asymptotic growth of the longest edge in an optimal solution to a MTSP. We will also progress on the persistently open question about the existence of a central limit theorem for the MTSP.

Motivations and Significance

Following Karp's paper [5], and in addition to Haimovich and Rinnooy Kan [2], many results have appeared in the literature about the asymptotical optimality (with probability one, in probability, and/or in expectation) of heuristics for various problems in the area of routing and location theory. For the practitioner, it would be interesting to know whether an asymptotically optimal heuristic is really applicable for realistic problem sizes, or whether its asymptotical behavior is only of theoretical importance.

A mandatory first step toward this general problem is to evaluate, for the optimal solutions, the error one makes by using an asymptotically valid formula when dealing with a finite size problem. Note that this second issue has also practical applications in its own, such as in strategic planning. In any case the fundamental starting problems in this area seem to be the analysis of rates of convergence and the question of the existence of central limit theorems.

References

[1] J. Beardwood, J. Halton and J. Hammersley (1959) "The Shortest Path Through Many Points", *Proc. Camb. Phil. Soc.* 55, 299–327.

[2] M. Haimovich and A. Rinnooy Kan (1958) "Bounds and Heuristics for Capacitated Routing Problems", *Math. Oper. Res.* 10, 527–542.

[3] P. Jaillet (1990) "Analysis of Probabilistic Combinatorial Optimization Problems in Euclidean Spaces". Submitted in revised form to *Math. Oper. Res.*

[4] P. Jaillet (1990) "Rates of Convergence for Quasi-Additive Smooth Euclidean Functionals and Application to Combinatorial Optimization Problems". Submitted in revised form to *Math. Oper. Res.*

[5] R. Karp (1977) "Probabilistic Analysis of Partitioning Algorithms for the Traveling Salesman Problem in the Plane", *Math. Oper. Res.* 2, 209–224.

[6] W. Rhee and M. Talagrand (1988) "A Sharp Deviation Inequality for the Stochastic Traveling Salesman Problem", *Ann. Proba.* 17, 1–8.

[7] J. Steele (1981a) "Subadditive Euclidean Functionals and Nonlinear Growth in Geometric Probability", *Ann. Proba.* 9, 365–376.

[8] J. Steele (1988) "Growth Rates of Euclidean Minimal Spanning Trees with Power Weighted Edges", *Ann. Proba.* 16, 1767–1787.

Stochastic and Dynamic Vehicle Routing in the Euclidean Plane

Dimitris J. Bertsimas* and Garrett van Ryzin**

* Sloan School of Management, RM E53–359, Massachusetts Institute of Technology,
Cambridge, MA 02139, U.S.A.
**Operations Research Center, Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.

1. Introduction

The classical approach to vehicle routing problems is to view them as static and deterministic. A set of known customer locations defines an instance, and the objective is to visit customers so as to minimize the total travel cost subject to certain constraints (e.g. a limit on vehicle capacity). These classical problems have generated significant research interest over the years resulting in major contributions in the areas of combinatorial optimization, the analysis of heuristics and complexity theory. However, as models for the type of vehicle routing problems encountered in practice they are not always appropriate. Many real-life problems involve considerable uncertainty in the problems data. For example, locations may be known only probabilistically in advance and the demands they place on vehicle capacity may be random. In addition, requests for service often arrive sequentially in time, and again these arrivals epochs may be stochastic. Finally, the objective of minimizing travel distance is not necessarily paramount; in a dynamic setting, the delivery time (wait for service) is often a more appropriate objective.

As a canonical example of an application with these characteristics, consider the following utility repair problem: A utility firm (electric, gas, water and sewer, highway, etc.) is responsible for the maintenance of a large, geographically dispersed facilities network. The network is subject to failures which occur randomly both in time and space (location). The firm operates a fleet of repair vehicles which are dispatched from a depot to respond to failures. The vehicle crews spend a random amount of time servicing each failure before they are free to move on to the next failure. The firm would like to operate its fleet in a way that minimizes the average downtime due to failures.

There are many closely related problems to this canonical example that arise in practice. For example, consider a film that delivers a product from a central depot to customers based on orders that arrive in real-time. Orders are queued and delivered by a fleet of vehicles with the objective of minimizing the average wait for delivery. In still other applications such as emergency service system operations (e.g. fire, ambulance and police) or courier services similar characteristics and criteria apply.

Motivated by these application areas, we propose and analyze generic mathematical models which we call the dynamic travelling repairman problem (DTRP). The DTRP is defined as follows: a region A contains one or many vehicles (servers) that travel at velocity v. Demands for service arrive according to a Poisson process with rate λ and, upon arrival, are independently and uniformly assigned a location within A. The problem is to find a policy for servicing the demands that minimizes the average system time, which includes both the waiting time in queue and the on-site service time. We also consider variations of the problem with many vehicles, capacity constraints on the vehicle and non-uniform demand.

The DTRP has several important characteristics:

1. The objective is to minimize waiting time not travel cost.
2. Information about future demands is stochastic.
3. The demands vary over time (i.e. they are dynamic).
4. Policies have to be implemented in real time.
5. The problem involves queueing phenomena.

In general, little is known about dynamic vehicle routing problems. In the next sections, we review our results.

2. Single Server DTRP Results and Policies

First, in the light traffic case ($\lambda \to 0$), we show the following policy is optimal:

- **The Stochastic Queue Median (SQM) Policy:** Locate the server at the median, x^*, of A and serve customers in FCFS order, returning to the median after each service.

In this case we showed that the expected optimal system time, T^* satisfies

$$T^* \sim \frac{E[\|X - x^*\|]}{v} + \bar{s} \quad \text{as} \quad \lambda \to 0,$$

where X is a uniform location in A. (Note that the first term above is simply the expected travel time to the median). However, this policy quickly becomes unstable as the traffic intensity increases.

In heavy traffic, we discover a quite different and unexpected behavior. If we let $\rho \equiv \lambda \bar{s}$ denote the fraction of time the vehicle spends in on-site service, then we showed that policies exist that have finite system times for all $\rho < 1$. This is surprising in that the condition is completely independent of the service region size and shape. It is also the mildest stability restriction one could hope for.

We then show that there exists a constant γ such that

$$T^* \geq \gamma^2 \frac{\lambda A}{v^2(1-\rho)^2} - \frac{1-2\rho}{2\lambda}.$$

Note that this grows like $(1-\rho)^{-2}$. Thus, though the stability condition is similar to a traditional queue, the system time increases more rapidly as $\rho \to 1$.

We construct several policies that have finite system times for all $\rho < 1$. In addition, we show that these policies have the same asymptotic behavior, namely

$$T \sim \gamma_\mu^2 \frac{\lambda A}{v^2(1-\rho)^2} \quad \text{as} \quad \rho \to 1,$$

where the constant γ_μ demands only on the policy μ. Hence, by comparing this to the lower bound above, we see that the policies have a constant factor performance guarantee in heavy traffic relative to the optimum value T^*. The policies and their associated constants are given below. When an exact analysis fails or only bounds were possible, empirical values for the constants were obtained via simulation.

- **The Partition (PART) Policy:** (This policy applies to the case where A is a square.) Divide the region into n equal subregions which are served sequentially such that each subregion is adjacent to the next subregion in the sequence except, perhaps, for the last one. Within each subregion, service demands in FCFS order until no more demands are left (exhaustive service). Then move on to the next subregion in the sequence. Optimize over n.
  Constant value: $\gamma_{PART} = \sqrt{2c_1} \approx 1.02$, ($c_1$ is the expected distance between two uniform points in the unit square.)
- **The Traveling Salesman (TSP) Policy:** As demands arrive, form them into sets of size n. When all n demands have arrived, consider it the arrival of a set. Service sets in FCFS order by forming a TSP tour on the set of demands. Optimize over n.
  Constant value: $\gamma_{TSP} = \beta_{TSP} \approx 0.72$, ($\beta_{TSP}$ is the asymptotic TSP constant.)
- **The Space Filling Curve (SFC) Policy:** A space filling curve is a mapping of locations from the unit square to positions *(preimages)* on the interval [0,1]. Using a real-time sorted list, maintain the preimages of all outstanding demands in the system. Visit demands according to the order in which they are encountered in continuous sweeps of the interval [0,1].
  Provable Bound on Constant Value: $\gamma_{SFC} \leq 2$, Empirical Value: $\gamma_{SFC} \approx 0.66$.

It is remarkable that this diverse collection of policies have identical asymptotic behavior. Moreover, this behavior has a particularly simple form that clearly shows how the average system time behaves as a function of vehicle velocity, service region size, on-site service statistics and traffic intensity. As mentioned, the behavior is distinctly different from that found in traditional queueing systems. In particular, it is proportional to $(1-\rho)^{-2}$ rather than $(1-\rho)^{-1}$ and also does not depend on the on-site service time variability (at least in its leading behavior).

3. The Many Vehicle Capacitated Case

As satisfying as these results are, the model of a single uncapacitated vehicle is somewhat unrealistic for most practical purposes. Therefore, we were motivated to expand the analysis to more realistic configurations. We thus extend our results to the case where the region A is serviced by a homogeneous fleet of m vehicles operating out of a set D of $|D| = m$ depots, where each vehicle is restricted to visiting at most q customers before returning to is respective depot. (The depot locations need not be distinct.) We show that the minimum expected system time, T^*, in this case has the following lower bound:

$$T^* \geq \gamma'^2 \frac{\lambda A\left(1 - \frac{1}{q^2}\right)}{m^2 v^2 \left(1 - \rho - \frac{2\lambda \bar{r}}{mqv}\right)^2} - \frac{m(1-2\rho)}{2\lambda},$$

where $\rho \equiv \lambda \bar{s}/m$, $\bar{r}$ denotes the expected distance from a uniform location in A to the closest point in D and γ' is a numerical constant. Note that for the case $q \to \infty$ and $m = 1$, this reduces to our earlier result.

When $m > 1$ and $q \to \infty$, we show that policies with the same constant factor performance guarantee as in the single server case can be constructed by simply partitioning A into m equal subregions and serving each one independently using one of the single server policies mentioned above.

When q is finite, the above expression provides some intuitively satisfying insights. For example, consider the case where $m = 1$ and $q < \infty$. Then the above bound implies a stability condition of

$$\rho + \frac{2\lambda\bar{r}}{vq} < 1.$$

We show this condition is also sufficient by constructing policies with this same stability criterion and asymptotic behavior. Observe that this condition is no longer independent of the service region geometry in this case because of the presence of $\bar{r}$; however, for $q \to \infty$ the dependence vanishes.

The second term in the stability condition has the interpretation of a *radial collection cost* in the sense of Haimovich and Rinnooy Kan. That is, $2\bar{r}/v$ is essentially the average time required to reach a set of q customers from the nearest depot (the radial cost). Dividing by q gives the average radial cost per customer, and hence multiplying by λ we obtain the fraction of time the server spends in radial travel. The above condition says that as long as this fraction plus the fraction of time spent on-site is less than one, the system will be stable. Furthermore, the waiting time grows like the inverse square of this stability difference just as it does in the uncapacitated case. Thus, we see the crucial role the average distance $\bar{r}$ plays in the system's stability. Indeed, we prove that if one has the option of locating the depot anywhere within A, then minimizing $\bar{r}$ (i.e. locating the depot at the median) is always optimal in heavy traffic.

For $m > 1$ and q finite, we construct policies that have system times within a constant factor of the optimum for several cases. In the case where all m vehicles are based out of the same depot, we show that a policy based on subdividing the region into squares, forming tours of q customers within each square and then serving tours in FCFS order has a constant factor guarantee. When there are k depots, this policy can be applied provided certain symmetry conditions hold. Specifically, the k Voronoi cells must be identical and there must be $m = kp$ vehicles for some integer p. Also, when k is large and the depots occupy the k-median locations, then the Voronoi cells are approximately identical and similar policies can again be constructed.

For completeness, we also discuss extensions to regions in higher dimensions and to locational distributions that are not uniform. In addition, we use our results to formulate and analyze several strategic planning problems such as optimization of travel and waiting cost, optimal fleet sizing and optimal districting.

4. Importance of the Contribution in Transportation

As mentioned before, real-world vehicle routing problems are characterized by time-varying and uncertain demands, open-ended planning horizons, congestion and a variety of performance criteria that include waiting time along with delivery cost. Yet, classical vehicle routing problems are static, assume perfect knowledge of demands, ignore queueing effects and typically only consider travel costs. Procedures have been developed to extend these classical models to more dynamic problems using, for example, rolling horizon approximations; but these offer little insight into the effect of time-variability and uncertainty on system performance. Using techniques from queueing theory, probabilistic analysis, geometrical probability and simulation, our paper is, to the best of our knowledge, the first to characterize the waiting time behavior of dynamic/stochastic demands serviced by a set of vehicles traveling in the Euclidean plane. The results give insight to the behavior of the system and show that the waiting time behavior is distinctly different than might be expected from classical queueing theory results. We characterize the performance of several diverse policies and compare their performance to lower bounds on the optimal waiting time. Policies that are provably good in both light and heavy traffic are proposed.

We believe that the model and the results will, in the short term, be used for transportation planning problems where congestion effects dominate, and, in the long term, promote more intense research into the interface between vehicle routing and queueing theory.

Session 23
Traffic Management – Gestion de la circulation

Président / Chairperson:	Nathan Gartner
Date:	Tuesday, June 11 Mardi, 11 juin
Heure / Time:	10:30
Salle / Room:	Château Frontenac Salon Frontenac

Smart Traffic Signals

Richard C. Larson* and Karla V. Ballman**

Operations Research Center, Room E40–164, Massachusetts Institute of Technology,
Cambridge, MA 02139, U.S.A.

Our focus in on **smart traffic signals**, made intelligent by sensors reporting real time traffic conditions at and near each signalized intersection. Using available technologies for both isolated and coupled intersections, our research hypothesis is that national investment in smart traffic lights will provide a dramatic ROI (Return On Investment), paying back within one year when measured by the following three performance measures: (1) the time delay experienced by the occupant(s) of a vehicle, (2) the amount of fuel consumed, and (3) the amounts of different pollutants emitted. It may not be possible to develop a system that will optimize all three measures simultaneously, so we seek Pareto optimal solutions. Various weighted performance functions are developed and analyzed for their reasonableness with regard to the numerous groups affected by signals. These groups include policy makers, commuters, commercial vehicle drivers, companies that rely heavily on road transportation and environmental groups. In addition to determining which performance measures are reasonable for each specific group, we identify potential conflicts of interest between the groups. Lastly, we choose a performance measure that seems to be the most suitable to the general public.

Using the performance measure we developed, we use our simulations and models to compare systems in use today with a hypothetical system operating under perfect information. From this we obtain an approximation to the value of the perfect information in comparison to current control systems. Next, we calculate the savings associated with upgrading all existing systems to the best technology currently available. We then improve an algorithm that has been shown to work better than signals operating under vehicle actuated control. This is a real-time adaptive control algorithm based upon dynamic programming. Using the aforementioned methods, we evaluate the algorithm's performance and then proceed with a cost-benefit analysis for implementation of such systems nationwide. Lastly, we explore possibilities for other adaptive control algorithms not based on dynamic programming but rather on a more continuous flow of time. We present a comparison of all the above systems, the cost of operating each system, and discuss our discoveries.

We also extend the existing real-time adaptive, dynamic programming based control algorithm to a network of coordinated signals. We compare the system-wide performance measures under the adaptive system to that of current coordinated network systems and report our results.

Since we have demonstrated there are savings associated with upgrading the nation's traffic signals to the best technological systems currently available, and that there are even more savings associated with newer adaptive systems, we attempt to identify the necessary procedures that would lead to implementation of the improved strategies.

Contribution to the Field

We view our work as a necessary intermediate step between the traffic control systems currently employed today and the intelligent vehicle/highway systems (IVHS) proposed for a future. The technology exists to implement the real-time adaptive control systems today and we have demonstrated the immediate savings that could be realized from implementing such systems. Furthermore, it will be necessary to implement such traffic control systems for the intelligent vehicle highway systems of the future. Our work can be viewed as an initial phase of IVHS.

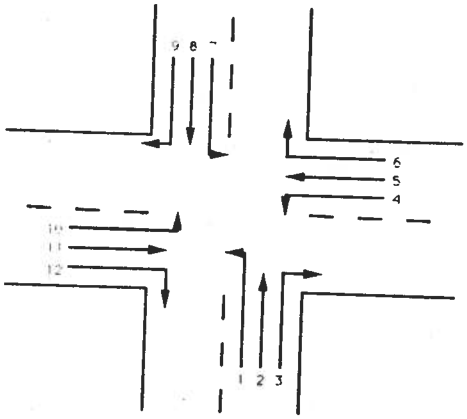
A Fast Algorithm for Signal Setting at Traffic Junctions

Christopher A. Moyer

Academic Faculty of Management Sciences, The Ohio State University, 301 Hagerty Hall,
1775 College Road, Columbus, OH 43210–1399, U.S.A.

In this paper a method for setting traffic signals at individual signalized junctions is presented. Setting of traffic signals involves determining the green time for each stream of traffic and the sequencing of the green signals. Typically the objective is to minimize the cycle time or to maximize a junction capacity factor. Usually one fixes one and optimizes the other.

Past solution techniques for this problem fall into two classes. The first includes methods which assume that the scheduling of green times, expressed by a stage matrix, is known, and therefore cannot guarantee a global optimum. Subsequently, these models were generalized to include the determination of the stage matrix. These more general models were shown to yield good but not necessarily optimal solutions. The second and more recent class of approaches to the problem is to treat it as a binary-mixed-integer-linear program (BMILP). By itself, this approach can only provide solutions to small problems within a reasonable amount of time. The diagram below shows a typical intersection with twelve traffic streams. The number of integer variables in BMILP is on the order of the number of streams squared. The junction illustrated below would have 144 integer precedence variables and over 900 constraints in BMILP. And it is very easy to imagine junctions much larger than the one listed.



A recent paper by Cantarella and Improta (*Transportation Research B*, vol. 22) uses a graph theory solution procedure in conjunction with BMILP. This method assumes knowledge of all the cliques in an associated incompatibility graph. Determining all the cliques in a graph is an enormously time consuming task, as a graph can have an exponential number of cliques. The method proposed by Cantarella and Improta also assumes that there are no odd holes in the incompatibility graph. It is easy, however, to imagine a junction where this assumption does not hold. If there are any odd holes their procedure cannot guarantee an optimal solution.

The solution method we propose makes none of the limiting assumptions of these past approaches. Further, our solution procedure does not require knowledge of all the cliques in the associated incompatibility graph. Our procedure involves solving a series of constrained shortest path problems, which can be done quickly. Speed of solution is essential for two reasons:

1. The problem needs to be solved a number of times so as to sequence lights at different times of the day (i.e. rush hour).
2. The decision maker will be interested in studying the trade-off between cycle time and volume of traffic flow.

The solution method we propose first considers the binary-mixed-integer-linear program discussed in the most recent papers. We set the cycle time and seek to optimize the junction capacity factor. The cycle time can later be varied parametrically, allowing the decision maker to consider additional options. Having set the cycle time, we then fix the green times for each stream of traffic within the constraints of the original program. The primal objective function involves only constants and variables we have fixed, so we now have to verify that a feasible solution exists for the cycle time and green times we have set. From the Primal-Dual Theorem of Linear Programming, to establish feasibility for the primal we need only to verify that the dual is bounded. Having reduced the model we note that the dual has a structure very similar to the classical shortest path problem. Specifically, the dual of the reduced model is a constrained shortest path problem where some pairs of arcs are restricted. The restrictions take the form: if one of the pair has flow, the other cannot. The constrained shortest path problem, which solves quickly, provides verification of feasibility.

Once we have determined a primal feasible solution, we can parametrically vary the green times. The number of break points to be examined in this search grows linearly with the number of traffic streams. Specifically, the maximum number of break points is the number of streams minus one. This allows us to maximize the junction capacity factor for our fixed cycle time. We can then vary the cycle time for the junction and resolve the model. The decision maker is then provided with a series of optimal capacity factors based on the different cycle times. This provides the decision maker with additional flexibility if factors outside the model influence the sequencing and timing of signals at a particular junction.

We test our approach extensively on various traffic junctions. Computational effort and other observations from the model are provided. The importance of developing fast solution procedures extends beyond the problem discussed here. Although typical traffic junctions have a relatively small number of streams, solution procedures similar to the one discussed here could be developed for other, larger problems involving the scheduling of incompatible flows through a junction. The procedure discussed here could also be used as a basis for the study of multiple traffic junctions, where each junction must be analyzed individually and with respect to the other junctions being studied.

A Branch-and-Bound Algorithm for the Traffic Signal Synchronization Problem with Variable Speed

Paul Mireault

Département de méthodes quantitatives et systèmes d'information, Ecole des H.E.C. and Centre de recherche sur les transports, Université de Montréal, C.P. 6128, succursale A, Montréal, Canada H3C 3J7

Traffic signal synchronization is a concern of traffic engineers because, even though they help prevent accidents, they can have undesirable side-effects such as increasing the number of stops a vehicle has to make, which in turn causes energy loss, excessive wear and tear on vehicles and driver aggravation. One of the most popular methods of synchronizing traffic signals is the progression or the green band method. The objective of this method is to let a group of vehicles, called a platoon, travel from one end of an artery to another without being stopped by a red light. Even though this method does not take into account traffic density and interference from turning vehicles, it gives good results.

Solution methods have ranged from heuristics using a scale replica of the artery with threads and sliding strips of paper with red markings representing the red signals (Petterman [1947]), to exact algorithms using a mixed-integer linear programming formulation (Little [1966]). Since then, Little [1980] developed MAXBAND, a FORTRAN program that is distributed by the U.S. D.O.T., and Mireault [1988] used Benders' decomposition method to help solve bigger problems faster.

The subproblem used in Benders' decomposition method has a network flow structure that can be taken advantage of: it can be viewed as a network flow problem where we want to maximize the "common flow" of some arcs. Here is a brief description of the subproblem structure:

$$\text{Maximize} \quad b$$

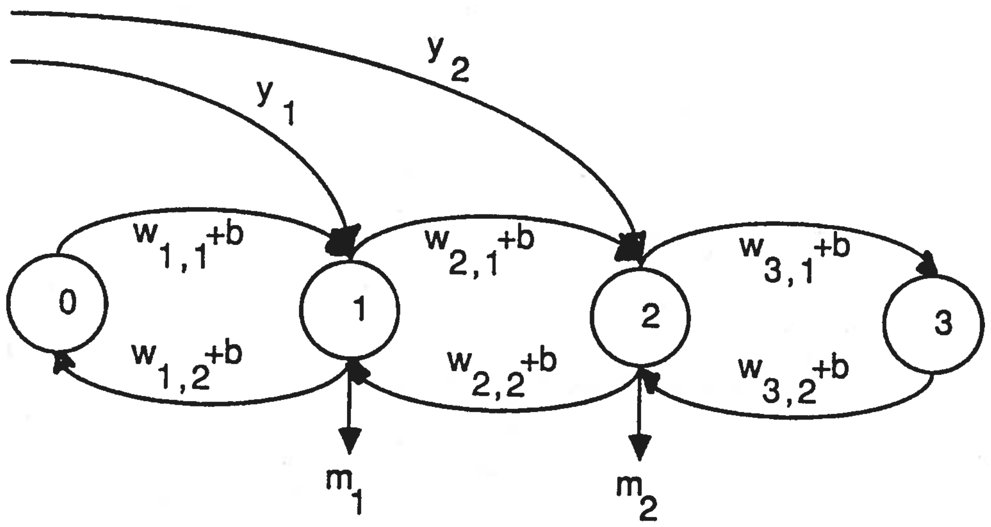
subject to

$$(w_{i,1} + w_{i+1,2}) - (w_{i+1,1} + w_{i,2}) + y_i = m_i, \quad i = 1, \ldots, n-1$$

$$0 \leq w_{i,d} + b \leq g_{i,d}, \qquad i = 1, \ldots, n,\ d = 1, 2$$

$$y_i^- \leq y_i \leq y_i^+, \qquad i = 1, \ldots, n-1$$

Constraints (1) can be viewed as a flow conservation equation and constraints (2) and (3) are the flow bounds. The following figure illustrates the problem:



In this presentation, I will start the problem formulation. Then, I will present an $O(n^2)$ algorithm to solve the subproblem. Finally, I will show how to solve the original problem, where the m_i's are variables, with a branch-and-bound algorithm that takes advantage of the structure of the subproblem's algorithm. I will compare the timings of this branch-and-bound algorithm with the timings of the ordinary mixed-integer linear programming algorithm.

INDEX

Authors