Mass Transport Vehicle Routing Problem (MTVRP) and the Associated Network Design Problem (MTNDP)

R. Jayakrishnan

Department of Civil Engineering and Institute of Transportation Studies

University of California, Irvine, CA. 92967, U.S.A.

Ph: (949) 824 2172, Fax: (949) 824 8385

Email: rjayakri@uci.edu

Laia Pagès

Department of Civil Engineering and Institute of Transportation Studies

University of California, Irvine, CA. 92967, U.S.A.

Ph: (949) 824 5623, Fax: (949) 824 8385

Email: lpages@uci.edu

Mass Transport Vehicle Routing Problem (MTVRP) and the Associated Network Design Problem (MTNDP)

ABSTRACT

This research studies a new class of dynamic problem MTVRP where n vehicles are routed in real time in a fast varying environment to pickup and deliver m passengers when both n and m are big. The problem is very relevant to future transportation options involving large scale real-time routing of shared-ride fleet transit vehicles. Traditionally, dynamic routing solutions were found as static approximations for smaller-scale problems or using local heuristics for the larger-scale ones. Generally heuristics used for these types of problems do not consider global optimality. A hierarchical method to solve the MTVRP in three different stages has been developed. Within the optimization process, a particular case of Network Design Problem (NDP) is solved. This paper introduces MTVRP and presents a scheme to solve it. Then, it describes the associated Mass Transport Network Design Problem (MTNDP) and solves the problem. The computational complexities as well as the results are compared.

1- DESCRIPTION OF MTVRP

Complex systems of fleet management, fleet control and vehicle routing problems (VRP) have widely been treated, but sometimes the problems treated have no application in the real world because they are too complex an have to be oversimplified in order to be studied, as stated by (Bodin 1990) and (Chen et al. 2003). In our case, the problem treated has no restrictions in terms of the formulation itself and can be applicable to a wide range of different areas.

A well-known problem with parallels to MTVRP is the dial-a-ride problem. (Savelsberg 1995) reviewed the various approaches used in the research community and they divide the pickup and delivery problems into static and dynamic, with and without time windows. But in any case, demand is introduced in real time while vehicles are already being routed; in other words, the pattern of demand is fully defined before scheduling takes place. Nevertheless, the assumption that all trips are known in advance may be reasonable in some situations. For example, freight companies face the problem of matching a fleet of vehicles to cur-

rent and future demand, but in these cases demand, fleet size, service times, travel times, and travel costs are known a priori. On the other hand, in a dynamic approach such as MTVRP nothing is known about the demand a priori, therefore the system should be able to respond to the demand as it appears in real time.

Approaches that rely on vehicle routes found as in a static VRP problem (for goods delivery) and modifying it for real-time passenger demand have been proposed in the past. A study by (Jaw et al.1986) used a sequential insertion heuristic algorithm for the dial-a-ride problem. The work of (Horn 2001) is also of particular interest to our research. This study models a multimodal system comprised of taxis, demand-responsive services and conventional timetabled services, such as buses or trains.

Recent research by the authors (Cortes, 2003) has developed a new demand-responsive transit system design called High Coverage Point to Point Transit (HCPPT). The algorithms therein tackled the problem primarily in a decomposed manner, with local control using heuristic operational rules containing notions of optimal directions. A question arises in Cortes' HCPPT model, on how much the system's performance can be improved by changing the vehicle deployment patterns.   This aspect was pre-specified in that work (the vehicles were assigned to operate in areas). Whether a periodic re-optimization using an appropriate class of simplified optimal formulations would improve the HCPPT performance motivated this research.   A second issue, perhaps more important involved the transfer points in the system, which were also pres-specified in that research. Our further research quickly arrived at the need for developing a class of algorithms that are fundamental to passenger transport with a large fleet of transit vehicles which are not assigned to lines or schedules.

2- METHODOLOGY

Generally heuristics used for dynamic routing problems do not consider specifically global optimality. A procedure to solve the "MTVRP" through three different stages, which seeks optimality by performing a hierarchical optimization is presented in this section.

But the fact that the routing decisions are taken in real time does not allow the problem to have a closed form. Then, the optimization problem treated cannot be solved with a normal classical optimal control formulation because the formulation does not explicitly enumerate search directions at any of the solution points. . Thus a question arises on how well the optimization schemes perform. Simulation of the optimization schemes seems to be the only viable way to study this. Due to the dimensions of the problem and the nonlinearities in the costs (passenger costs in vehicle rerouting depends on the load in the vehicles), microsimulation is the plausible tool within this analysis.

For the sake of brevity, it is not possible to enter into the details of each of the three stages, but it is important to give an overview of each of them in order to understand the whole process. Therefore, a brief introduction of each of the stages is presented. Then, since the part having more interest from an Operations Research point of view is the MTNDP, this paper focuses on this part, presenting the details of the formulation and describing the solution scheme. Finally, solutions are presented for several examples and conclusions are extracted.

- STAGE 1- NETWORK AGGREAGTION In the first stage of this procedure the demand is grouped into zones in order to diminish the complexity of the network. Each zone has associated with it: a) One Centroid (similar to Centroids in planning models) and b) groups of passengers willing to go from its origins to its destinations. The topological form and distribution of these zones is very important and can be crucial in the quality of the solution. As noted above, this part of the procedure is not explained in-depth in this paper.

- STAGE 2- MTNDP - Solution of the "Mass Transport Network Design Problem" Solution of an acceptable static problem, which is a simplification the proposed dynamic one in the aggregated network described above. This paper focuses on this stage of the procedure and it is further explained below.

- STAGE 3- LMTVRP Solution of the "Local Mass Transport Vehicle Routing Problem" The solution to MTNDP is introduced under the form of bus routes and frequencies or "density of buses" for each path. Once the buses have an initial position, a local routing scheme takes place for each zone (taken independently). The number of buses being low, known vehicle routing heuristics are used to route the vehicles. The cost function coded is similar to the one used by (Cortes, 2003) where the cumulative travel time of passengers in the system is considered. Again, this part of the process is not presented in detail in this paper.

3- MASS TRANSPORT NETWORK DESIGN PROBLEM

3.1- FORMULATION

Consider a network represented by the directed graph G (N, A), with node set N and arc set A. The arcs here can be viewed as "global vehicle corridors" between nodes which represent "areas" (or even transfer hubs) in a large urban area. The intent is to find the "flow" (or frequency) of vehicle movement on these corridors, and to find how many vehicle travel on paths that go over such "nodes", so that a global estimate can be made for mass transport vehicle movements. Note that the actual vehicle movements may involve local rerouting of these vehicles on actual network links for passenger and pick-up and delivery, as handled by the stage-3 of the scheme. The global path and flow solution and the associated vehicle deploy-

ment (mass transport network design) can then be used as the starting solution or the solution that set the global parameters for the LMTVRP problem (stage-3). One can also view MTNDP as solving the trunk route movement problem, if these vehicles are to have non-reroutable trunk route travel and re-routable local travel as in the HCPPT problem (Cortes, 2003). The following discussion now focuses on these global "paths" and vehicle flows on them. Though we use the term "stops" as in the standard transit network design problem, these are not actual bus stops but are rather areas (represented by nodes) where vehicles enter the local streets or terminals rather than travel through as in express transit corridors. Thus the solutions would tell us how many vehicles travel from which areas in a large urban network to which other areas, at what type of frequency. The design thus provides us vehicle "routes" which can be viewed similar to bus lines or services in a bus network design problem, but are referring to global vehicle route corridors rather than bus lines or services with set schedules and frequencies.

Consider a set of the paths (p_{ijk} being the kth path starting at node i and finishing at node j). Some of the routes might have sub routes, which have both its origin and destination coinciding with intermediate vehicle stops of the longer vehicle routes. Let us call the long route "parent" and the sub route "child". For the sake of clarity, Figure 2 draws a simplified example of paths p_{ijk}, $p_{ijk'}$ $p_{ijk''}$ starting at node i and ending at node j and paths p_{lmn} and $p_{lmn'}$ starting at node l and ending at node m. In this example, p_{ijk} is a parent of p_{lmn}.

There are three sets of constraints: (1.1) continuity constraint; (1.2) link capacity constraint and (1.3) total fleet constraint (vehicle availability). The formulation for MTNDP is the following:

$$\textit{Minimize} \quad \sum_i \sum_j \sum_k \sum_l \sum_m \sum_n f_{ijk}^{lm} x_{lm} (tt_{ijk}^{lmn} + wt_{ijk}^{lmn}) \delta_{ijk}^{lmn} \qquad (1)$$

Subject to:

$$\sum_i \sum_j \sum_k f_{ijk}^{lm} = 1 \qquad\qquad \forall l, m \qquad\qquad (1.1)$$

$$\sum_l \sum_m f_{ijk}^{lm} x_{lm} \phi_{lmn}^a \delta_{ijk}^{lmn} \le Cr_{ijk} \qquad \forall i, j, k \qquad\qquad (1.2)$$

$$\sum_i \sum_j \sum_k tt_{ijk}^{ijk} r_{ijk} \le N \qquad\qquad\qquad (1.3)$$

p_{ijk} = kth path between node i and node j

r_{ijk} = rate of vehicles on path p_{ijk} (vehicles/unit time)

x_{ij} = rate of demand from node i to node j (passengers/unit time)

tt_{ijk}^{lmn} = in-vehicle travel time from l to m using p_{lmn} , child of p_{ijk} (unit time)

wt_{ijk}^{lmn} = waiting time from l to m using p_{lmn} , child of p_{ijk} (unit time)

f_{ijk}^{lm} = fraction of demand x_{lm} traveling on p_{ijk} , parent of p_{lmn}

δ_{ijk}^{lmn} = 1 if p_{lmn} is a child of p_{ijk} , = 0 otherwise

ϕ_{ijk}^{a} = 1 if p_{ijk} goes over link a, = 0 otherwise

C = Vehicle capacity (passengers / vehicle)

A close examination of the details would reveal some similarities between the MTNDP formulation and the well known Network Design Problem (NDP), studied by (Mandl 1980), (Hasselstrom 1981), (Ceder and Wilson 1986), (Baaj and Mahmassani 1994), (Bahnhart et al. 1998) and (Van Nes 2003), among others. But there exist differences between MTNDP and NDP that make our problem easier to solve. MTNDP is not a subset or a simplified version of NDP; rather, these are just two different problems. The primary difference is that NDP is geared towards finding a network design that is immediately followed by frequency setting and scheduling, whereas the MTNDP finds a solution at an abstract (or aggregate) level composed of transit "flows". Furthermore, the solution to MTNDP violates some of the basis for NDP, for example MTNDP identifies a situation without considering the return of the transit vehicles to the starting point of the vehicle route. The flow for each line "appears" at the origin of the line, and then travels along the route, picking up and dropping off passengers and later "disappears" at the destination (or terminus) of the line. Note that this would be impossible to implement in the real world, but its solution is what gives global optimality in the p[resent of local routing optimization.

In terms of the theoretical details, NDP presents a concave (Newell 1972) and multiobjective objective function arising from the fact that it tries to minimize both passenger and operator cost. Therefore, in the case of NDP, the tradeoffs among the conflicting objectives need to be addressed because it is multiobjective. But in MTNDP, only the passenger costs are minimized, therefore the objective function is simpler. Also, we do not accept transfers. Another difference is that the combinatorial explosion originating from the discrete nature of NDP is not present in MTNDP because its variables are continuous. Also, the network taken into account for MTNDP is an aggregated one; therefore the sizes of the networks needed are relatively small compared to the networks solved by researchers working on NDP (Barnhart et al 1988). Finally, our problem aims to create a transit system from scratch, whereas, in general, the most conventional

NDP schemes starts from an already created system adding new lines and/or changing the existing ones to optimize the system.

To further explain, imagine a "rate of passengers" willing to travel from the origins to the destinations. The solution for MTNDP is the vehicle "lines" and the associated frequencies to perform these operations with a minimum user cost. Therefore, our interest is to find the "lines" (which can be viewed as loose collection of fleet vehicles operating in some corridors) and vehicle frequencies between each OD pair, minimizing the travel time of users with two restrictions: a given fleet size (a), a maximum capacity for the vehicles (b). The length of the routes and the frequencies of the vehicles identify the vehicle deployment and trunk/express travel patterns in different areas of a real world network. In the next section, the steps performed to find a solution are described and an example is presented.

3.2- SOLUTION SCHEME

MTNDP is a minimum cost flow problem consisting of two levels. In the first level (we call it "Passenger Flow Problem" or PP) there are passenger flows, which are explicitly present in the objective function. In the second level (we call it "Transit Problem" or TP) there are vehicle flows (or "empty seats" flows). The iterative solution scheme is presented in Figure 1 and further explained below.
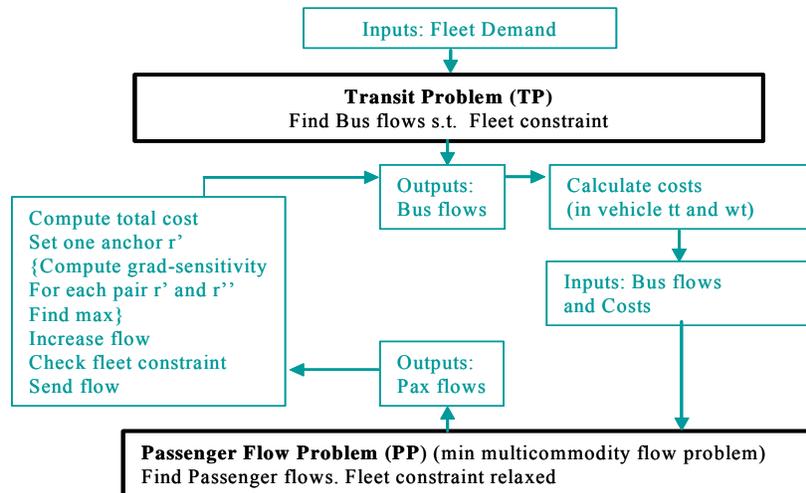

Fig. 1. Solution Scheme for MTNDP

Note that the TP is solved only once at the beginning of the procedure. Then, at each iteration a different PP is solved by changing the vehicle rates in the solution of the TP in the direction that decreases the MTNDP's objective function. The method first solves the TP heuristically. The heuristic consists on three parts.

1) First a "feasible initial solution" is found by setting each vehicle to its "minimum desired rate". From each node i to each other node j only the shortest path is considered and its rate is set equal to the rate of demand of passengers traveling from i to j, multiplied by the cost of traveling from i to j and divided by the capacity of vehicle units(the number of seats in each vehicle).   The solution obtained under the "minimum fleet needed" step is feasible because all the constraints are satisifed, but it is sub optimal because each passenger (nodal O-D pair) is using a line with the termini at the two nodes. On the other hand, under this solution, all vehicles are full all the time. Hence,the number of vehicles needed in this solution (let us call it F1) is expected to be lower than the optimal solution.

2) Secondly routes are integrated. For each arc of each parent route, the rate of vehicles for the parent route is set equal to the maximum of the sums of all the rates needed for all child routes. The operation is performed for all parent routes with a maximum preset number of links, starting by the parent route with the highest number of child routes. The procedure checks the fleet constraint at each increase of vehicle flows. This step of the process eliminates redundant lines and creates empty space in parent line's vehicles. Let us call the fleet size needed in this step of the process F2. In any case F1 < F2, but neither is larger than the maximum fleet size (N) fixed in the MTNDP formulation.

3) Finally, routes are truncated. Once rates for all parent routes are set, child routes are not needed anymore. But, if all child routes were cancelled, the resulting network would only be composed of high frequency parent routes and the fleet constraint might be violated. Equilibrium between parent routes and child routes is achieved here by canceling the child routes which have a lower rate than a predefined value and setting the rest of the child rates to a variable fraction of its previous rate. Let us call the fleet size used in this situation F3. Note that F3 > F1 and F3 < F2. Again, none of the fleet sizes are larger than the maximum fleet size N.

Once the above heuristic is completed, the PP problem can be optimized. The original network topology is changed to take into account the operational restrictions of the vehicles introduced in our model. First of all, transfers are not allowed. Secondly, we are routing vehicles as if they were taking part of a public transportation service line, therefore passengers who are traveling in a sub-path of any longer path can enter into a vehicle that is serving the bigger path but not vice versa. An "expanded" network depicting these phenomena is built, for use in the linear multicommodity flow optimization in the PP problem that follows . Each OD is represented by a super sink and a super source, corresponding to one commodity and each considered path is represented separately.

The TP solution allows the computation of the expected passenger waiting time for each line. A matrix with all the waiting times for each line is computed and is fed into PP, explained in the herewith.

As noted above, the solution for the TP through the heuristic described above is a set of routes and its associated frequencies. Note that, for a given set of vehicle rates, the waiting times for all routes are known and the PP associated to MTNDP becomes a linear multicommodity minimum flow cost problem, dealing only with a unique set of variables: passenger flows. Linear Multicommodity minimum flows problems have been solved in the past either with partitioning methods, resource directive decomposition or price directive decomposition. Lately other algorithms have been developed such as the interior point methods (Karmarkar 1984). According to (Kennington 1978), (Ali et al. 1980) and (Ahuja 1993) price directive decomposition methods perform better when a lot of commodities are considered; therefore a variant of this method has been implemented in this research.

Price directive decomposition eliminates the bundle constraints in the matrix by placing prices to these constraints and introducing them into the objective function. The resulting system is a relaxed "minimum flow cost problem" for each commodity. The problem has a dual variable for each arc (w_a) corresponding to the continuity constraint and a dual variable for each commodity (σ_k) corresponding to the link capacity constraint. Following this formulation, a pricing problem is solved at each step. If the complementary slackness conditions hold, then the solution is optimal and the procedure stops, if not, a new pricing problem is solved until an optimal solution is found. Interested readers on this method are directed to (Kennington and Helgason 1980) and (Kim and Barnhart 1999).

Once PP has been solved, we have an initial feasible solution. From this point, a particular implementation of a steepest descent method has been coded to optimize the problem. The essential idea behind our gradient-search is that each solution of TP gives one point in the solution space and the objective value for the MTNDP problem presented in section 3.1. The optimization finds the gradient using a perturbation of the solution and the global optimization proceeds using a steepest descent method. The reason why this procedure is selected is that the TP solution can be found relatively easily using the LLMCFP scheme. This explains the need for us to split the optimization problems to two subproblems, PP and TP. The procedure is as follows:

4- COMPUTATIONAL RESULTS

Seven different network sizes were studied. Table 1 reports the topology of the networks (number of links and number of commodities) and shows the computational time taken by CPLEX and by our algorithm to solve PP.

Then, the whole MTNDP was solved for two of the networks. Tables 2 and 3 report the objective function and the computational times taken to solve MTNDP (for different values of λ and number of iterations).

Table 1 Solutions for Passenger Problem in MTNDP

Links	Commodities	Solution time (sec)		Benefit (%)
		CPLEX	PP	
5	9	0.37	0.0001	0.03
18	225	1.22	0.03	2.46
28	1471	5.26	0.46	8.75
34	3346	20.56	2.23	10.85
44	19401	370	205	55.41
56	100466	930	500.8	53.85

It is difficult to compare the results attained by other researchers with our studies because the nature of the problem as well as the nature of the networks studied by others is different. But we have compared the computational time taken by our method with the computational time taken by CPLEX. We can conclude that the benefit of using our algorithm becomes bigger for larger networks.

Table 2 Solution values for MTNDP. 9-Commodity Example

Δr	0.1	0.1	0.1	0.1	0.1	0.5	0.5	0.5	0.5
ITER	2	15	25	45	75	2	5	7	10
MTNDP	4	57	89	132	173	4	7	10	11
OF	189.83	169.98	164.97	162	160.87	178.603	164.91	162.37	160.25

Table 3 Solution values for MTNDP. 24-Commodity Example

Δr	0.1	0.1	0.1	0.1	0.1	0.5	0.5	0.5	0.5
ITER	15	25	45	75	100	5	7	10	15
MTNDP	19	34	53	84	113	7	11	14	23
OF	588.61	588.55	588.42	588.35	588.35	588.55	588.49	588.4	588.4

As expected, the time needed in our method to find an accurate solution for MTNDP is much higher.

5- CONCLUSIONS

This paper identifies a problem called the "Mass Transport Vehicle Routing Problem" (MTVRP). The proposed solution scheme for MTVRP concerns the solution of an associated problem called "Mass Transport Network Design" (MTNDP). In this paper a description of MTNDP was presented and a solution for it was characterized. Then solutions to several examples are presented. It is difficult to compare the results obtained with our code and the results obtained with other authors because the examples and the nature of the problems treated are different, but for the cases treated so far, our code seems to obtain optimality faster than the solutions found using straight implementations with available software.

ACKNOWLEDGEMENTS

REFERENCES

Ahuja, R K, Magnanti, T L, Orlin, JB (1993) Network Flows. Prentice Hall, Englewood Cliffs, New Jersey

Ali A, Helgason R V, Kennington J L , Lall H , (1980) Computational comparison among three multicommodity network flow algorithms. Operations Research 28: 995–1000

Baaj, Mahmassani (1994) Hybrid Route Generation Heuristic Algorithm for the design of Transit Networks. Transportation Research vol 3, No. 1: 31-50

Bahnhart C, Jin H, Vance P (1998) Railroad Blocking: A network Design Application. Operations Research vol 48, No. 4: 603-614

Bodin L D (1990) Twenty years of routing and scheduling. Operations Research, 38: 571-579

Ceder, Wilson (1986) Bus Network Design. Transportation Research vol. 20B, 1986: 331-344

Cortes C. (2003) High-Coverage Point-to-Point Transit (HCPPT): A new design concept and simulation-evaluation of operational schemes for future technological deployment. Ph. D. thesis, University of California, Irvine

Xu H, Chen Z, Rajagopal S, Arunapuram S (2003) Solving a Practical Pick up and Delivery Problem. Transportation Scinece, vol 37, No. 3: 347-364

Hasselstrom (1981) Public Transportation Planning – A mathematical programming approach. PhD Dissertation. Department of Business Administration, University of Gothemburg, Sweden

Horn M E T (2001) Multi-modal and demand-responsive passenger transport systems: a modeling framework with embedded control systems. Transportation Research Part A vol. 36: 167-188

Jaw et al. (1986) A heuristic algorithm for the multi-vehicle advance request dial-a-ride problem with time windows. Transportation Research Part B vol. 20B No. 3: 243-257

Karmarkar N K (1984) A new polynomial time algorithm for linear programming. Combinatorica 4: 373–395

Kennington J L (1978) A Survey of Linear Cost Multicommodtiy Network Flows. Operations Research 26(2): 209–236

Kennington J L, Helgason RV (1980) Algorithms for network programming. John Wiley & Sons, New York

Kim, Barnhart C (1999) Multimodal Express Package Delivery: A service Network Design Application. Transportation Science Vol 33, No. 4: 391-407

Mandl C (1980) Evaluation and optimization of urban public transportation networks. European Journal of Operational Research vol 5: 396-404

Newell G (1979) Some issues related to the optimal design of bus routes. Transportation Science, vol.13: 20-35

Van Nes (2003) Multi user-class urban Transit Network Design. Paper presented at the Annual Meeting of the Transportation Research Board, Washington DC.