

# UC Berkeley

## Research Reports

### Title

Assessing the Role of AVL in Demand Responsive Transportation Systems

### Permalink

<https://escholarship.org/uc/item/6xg377xj>

### Authors

Gillen, David  
Raffailac, Julie

### Publication Date

2002-05-01

CALIFORNIA PATH PROGRAM  
INSTITUTE OF TRANSPORTATION STUDIES  
UNIVERSITY OF CALIFORNIA, BERKELEY

## **Assessing the Role of AVL in Demand Responsive Transportation Systems**

**David Gillen, Julie Raffailac**  
*University of California, Berkeley*

**California PATH Research Report  
UCB-ITS-PRR-2002-16**

This work was performed as part of the California PATH Program of the University of California, in cooperation with the State of California Business, Transportation, and Housing Agency, Department of Transportation; and the United States Department of Transportation, Federal Highway Administration.

The contents of this report reflect the views of the authors who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the State of California. This report does not constitute a standard, specification, or regulation.

Report for TO 4119

May 2002

ISSN 1055-1425

# **Assessing the Role of AVL in Demand Responsive Transportation Systems.**

David Gillen and Julie Raffailac  
Institute for Transportation Studies  
University of California  
Berkeley CA

## Introduction

Many-to-many demand responsive transportation systems consist of vehicles which take passengers from their origins to their destinations within a service area. In dial-a-vehicle systems, in order to circumvent the undesirable feature of taxicab systems, vehicles are allowed to deviate from their direct route to serve other passengers and the emphasis is on building efficient tours to increase vehicle productivity. This strategy increases riding times but also increases average occupancy and productivity of the vehicles, and hence decreases average waiting times.

A similar problem is faced by the recently developed 'webvan' food delivery service which takes orders for groceries over the internet and commits to delivery to the order's home (or specified address) within a given time frame. The difference is there is a single origin for a defined market area but multiple destinations. It is possible to organize dial-a-transit system in a similar manner but it is not clear how this form of delivery configuration would impact the productive use of the vehicles, drivers and other factors used.

The interest we have in this area is to develop a model that can be used to assess the improvement in productive efficiency or of consumer welfare with the use of ITS applications, such as AVL in providing this service. AVL provides information to both the service provider and the service consumer. The information on the supplier side allows the dispatcher to allocate vehicles to achieve some objective. In most cases this has been defined to minimize costs or maximize productivity. We find this to be a narrow definition, after all this is a service industry and meeting the needs of customers should be the objective. This can be defined as maximizing consumer welfare or utility. Perhaps a more reasonable objective and one that recognizes the scarce resource issue is to have an objective of maximizing welfare which means the sum of consumer and producer surplus. Such an objective function recognizes the separate roles of customers and suppliers and the trade-off of increasing costs and increasing quality. Using this approach we should be able to say something about optimal market size, vehicles per market and optimal vehicle size.<sup>1</sup>

Below we develop the beginnings of an algorithm that can be used to measure the contribution of AVL to both passenger (customer) welfare and supplier efficiency. The development can be considered in four stages. In the simplest case demand is fixed spatially and market size is fixed. For a given geographic area (so many city blocks) and for a given fixed known demand that is distributed in some way around this geographic area, and for a homogenous type of demand solve for the optimal routing and scheduling according to some objective function. The objective function may be economic welfare (sum of consumer plus producer surplus), to maximize consumer welfare, to minimize costs. All are possible candidates.

In Stage 2, we change one assumption and that is to add to a fixed known demand, a stochastic demand that is randomly distributed both spatially and in time. In this case we do not know when or from where some calls are going to come in from people who want service that day. Solve the same problem as in stage 1 but with the new demand assumption. The real question here is how much added capacity does the firm need to carry to satisfy a given level of service or to maximize the stated objective function. Stage 3, moves from homogeneous demand to two types of demand, one from able passengers and one from disabled passengers or perhaps passengers who take a

---

<sup>1</sup> In many respects these are precisely the problems faced by an airline in choosing routes, frequency and aircraft size.

small amount of extra time to load and unload. This is akin to the variable lot size problem in logistics in that the time taken to service the customer will vary. We solve the same problem as in stage 2 but with the new demand assumption.

In the final stage we define the optimal market size. This is not unlike the web-van problem. In their case they will serve customers and at some point they will expand their size, select a new warehouse site and serve the new customers from this new warehouse. A demand responsive transit firm has a similar problem because it must decide at what point it needs a new vehicle to service the new demand. One can consider that the assumption that has been relaxed in this fourth stage is the fixed geographic size of the market.

## Literature Review

A satisfactory analytical model of dial-a-vehicle operations has yet to be developed. This is perhaps because dial-a-vehicle routing algorithms have become so complicated that they seem to defy mathematical modeling. At present, simulation models or empirical models based on simulated or real data seem to be the only alternatives open to the planner.

Daganzo (1978) develops a simple analytic model to predict average waiting and riding times of many-to-many demand responsive transportation systems. In order to model different operating strategies corresponding to the operations of taxicab and dial-a-vehicle systems, 3 different routing algorithms are studied applying the same modelling technique used for the many-to-one case. The resulting model is simple enough to facilitate parametric analysis of these systems and model results agree well with the results of simulation.

A many-to-many dial-a-vehicle system can be visualized as a two-stage queueing network where vehicle perform service on requests. When the service request is made, it joins a queue of requests awaiting pick up in stage 1. The requests move from stage 1 to stage 2 when a vehicle arrives at the customer's origin and picks up the customer. During stage 2, the customer is in the queue of passengers riding on a given vehicle. The request leaves stage 2 when he is delivered at his destination.

The model evolves in the following way. Assume  $M$  vehicle over an approximately circular (or square) service area  $A$  with a request arrival rate  $\lambda$ . Further assume that O&D are uniformly and independently distributed over the service area. The routing algorithm is characterized as follows. When a vehicle has a choice of stops, it *selects the next nearest one* in order to reduce the number of stops per unit time which may be viewed as a means of increasing productivity. However, this strategy does not confine to studying a single routing algorithm.<sup>2</sup>

The arrival process is assumed to be continuous, deterministic, and constant. This means travel distances between stops are a deterministic (*decreasing*) function of the number of stops the vehicle can select from, and boarding and alighting times are also deterministic. The average

---

<sup>2</sup> Different algorithms can be obtained by defining a set of stops from which the choice can be made : Algorithm 1= after each stop, the vehicle is routed to the nearest feasible point (whether an O or a D). Algorithm 2= the vehicle alternates pick ups and drop offs (always selecting the closest feasible pick up or drop off). Algorithm 3= the vehicle collects a fixed number of passengers and then delivers them.

distance  $d_n$  between a random point and the nearest of  $n$  independent random points in the area  $A$  for circular or quasi-circular shape is :  $d_n=0,5\sqrt{(A/n)}$ .<sup>3</sup>

Daganzo explores four algorithms. Algorithm 1 attempts to build the most efficient tours at all times so the vehicle is equally likely to select an O or a D. This is accomplished by setting the number of requests waiting  $N$  equals the number of requests in the vehicle considered  $n$ .<sup>4</sup>

The arrival rate  $\lambda$  equals the rate at which vehicle pick up requests (rate at which any one given vehicle picks up requests  $\lambda_b$  times the number of vehicle  $M$ ) :  $\lambda=\lambda_b * M$ .  $\lambda_b$  is the inverse of the time between two successive vehicle pick up stops :  $\lambda_b=(b_1+b_2+2*d_{n+N}/v)^{-1}$  and  $b_1$  and  $b_2$  are the boarding and alighting times ( $b_1=b_2=0,5\text{min}$ ).  $v$  is the average vehicle speed ( $=0,25\text{mi/min}$ ). This formula implies vehicles are always able to pick up the customers they select; in effect ignoring the stochastic phenomena.

From this structure we can deduce the formula for  $N$ , and  $\lambda_{\max}=M/(b_1+b_2)$  which represents the theoretical capacity of the system under algorithm 1. Furthermore, we obtain the waiting times  $w=N/\lambda$ , the riding times  $u=n/\lambda_b=NM/\lambda$ , and the total time in the system  $t=w+u$ .

In algorithm 2 the vehicle keeps approximately a constant number of passengers on board at all times by altering pick ups and drop offs.  $n_i$  is the number of passengers in vehicle  $i$  after a pick up. And  $\lambda_b^{(i)}$  is the rate at which vehicle  $i$  picks up customers. For a vehicle system with all queues having at least one customer :  $\lambda_b^{(i)}=(b_1+b_2+(d_{n_i}+d_{n+N})/v)^{-1}$  (inverse of the sum of times for a pick up trip and a delivery trip), the equilibrium of the stage 1 queue leads to :  $\lambda=\sum_{i=1..M}\lambda_b^{(i)}$  For a given average number of persons in the vehicle  $n$  ( $=\sum_{i=1..M}n_i/M$ ), the values of  $n_i$  that minimize  $n$  and consequently minimize the total time spent by a customer in the system are  $n_i=n$ ,  $i=1..M$  The total delay is minimized when the number of customers in each vehicle is equal.

We find  $N$  (only for  $\lambda<\lambda_{\max}$ ) and the total time spent in the system  $t$ . The approximate number of customers in the system is :  $M(n-0,5+N)$ . The value of  $n$  that minimizes  $t$  is the optimal number of passengers we should keep on each vehicle (for this  $n$  the system is under capacity as long as  $\lambda<M/(b_1+b_2)$ ). Consequently, we can find the optimal  $N$  and the optimal delay  $t$ .

Algorithm 2 has a longer waiting to riding time ratio than algorithm 1 ; this results from forcing vehicles to deliver customers even though a pick up might be more efficient. In congested systems, algorithm 2 provides a lower total time in the system for large numbers of vehicles. Indeed, for large  $M$  the riding time with algorithm 1 is very large compared to the waiting time and it makes sense to trade off some waiting for riding time.<sup>5</sup> With this algorithm it is possible to select  $n$  to obtain a desired riding to waiting time ratio, or to minimize a linear combination of waiting and riding time.

Algorithm 3 is used to reduce the variance of the riding times. Vehicles are assumed to go through a cycle consisting of a collection phase, where exactly  $n$  passengers are picked up, and a

---

<sup>3</sup> In practical situations, the travel factor  $r$ , which captures the circuitry of the network, is introduced to take into account the non-Euclidean metric for real street networks.  $r=1,27$  for 2-directional grid networks in circular or square areas. The value of  $r$  for grid-like networks is approximately independent of trip distance. So, one can safely assume :  $d_n=r/2\sqrt{(A/n)}$ .

<sup>4</sup> This guarantees the equilibrium for the stage 2 queues,  $n=N$ .

<sup>5</sup> For  $n=1$ , this algorithm reproduces the operations of a taxicab company.

distribution phase where the passengers picked up are taken to their destinations. For  $n=1$ , algorithm 3=algorithm 2.

The model has the following features. The length of the cycle  $C$ , the collection phase  $G$ , and the distribution phase  $R$ , are expressed as functions of  $n$ . The rate at which passengers are picked up depends on the cycle time for each one of the vehicles, and at equilibrium :  $\lambda = \sum_{i=1..M} (n_i/C_i)$  with  $i$  for the  $i$ th vehicle.

The best level of service is achieved when all vehicle pick up the same number of requests,  $n_i=n$ . As a result :  $\lambda = Mn/C$ . We find  $N$ , the capacity  $\lambda_{max}$  and the waiting time  $w$ . It is assumed that the rate at which vehicles collect (distribute) passengers remains constant throughout the collection (distribution) phase. During a typical cycle, the number of passengers in the vehicle increases linearly to  $n$  and then decreases linearly to 0. In total, for all passengers, the riding time per cycle is equal to  $n \cdot C/2$  and on average the riding time per passenger is equal to  $C/2$ . There is an optimal value of  $n$ , that minimizes the total travel time  $t$ , which can be found by trial and errors.

Algorithm 3 tends to behave similarly to algorithm 2; both control the number of passengers on the vehicle and therefore both may be advantageous in similar instances. Algorithm 3 will in general be less efficient than algorithm 2 because for equal riding times algorithm 2 results in lower waiting times as it can process requests more rapidly.<sup>6</sup>

In our assessment of these models we see waiting to riding time ratios are very sensitive to the routing strategy used. However, total time is not as sensitive to the routing strategy and one can use algorithm 1 for most design purposes. Algorithms 2&3 can be more effective predictors of waiting to riding time ratios. The number of people in the vehicle is an exogenous variable which can be adjusted to reflect the conditions of the system being modeled by those algorithms 2&3. Furthermore, all the previous formulae for algorithms 2&3 are accurate for moderate demand levels (more than for algorithm1). Riding time is the most significant part of total delay

To take into account the stochastic pattern occurring for uncongested systems, we model the request arrival process as a time-homogeneous Poisson process and the service rates as mutually independent negative exponential variables independent of the arrival process - requests arrive at random, Poisson in time, uniform and independent in space. This paper deals with a simple model of many-to-many dial-a-vehicle systems that makes it possible to predict quickly and accurately average total time in the system. It is thus possible to design and evaluate planned dial-a-vehicle systems. It discusses the modeling framework and presents asymptotic formulae for 3 different algorithms ; it shows that formulae can be modified to capture stochastic phenomena and that those modified formulae are relevant with simulated results for both congested and uncongested conditions (algorithms 2&3 may not have to be corrected). A further study would be to explore the effects of correlated origins and destinations on the capacity and level of service that can be provided by the system, this is only relevant for large service areas.

Daganzo in a 1984 paper presents a preliminary study of the feasibility of checkpoint dial-a-ride systems. Their cost-effectiveness is compared to that of fixed route systems with no transfers and door-to-door dial-a-ride systems. A checkpoint dial-a-ride system combines characteristics of the other two. It resembles a door-to-door system because passengers are picked up and dropped off

---

<sup>6</sup> For our purposes algorithm 3 is important since it will tend to provide more reliable riding times as the maximum ride can never exceed  $C$ .

on demand at a finite number of locations near their trip ends, called checkpoints. These checkpoints are scattered over the area studied and their number affects the mode of operation considerably. The system can also resemble fixed route systems because patrons have to walk to and from the checkpoints.

The results are derived for a simple routing strategy, and involve some simplifications, which facilitate the comparisons. The goal is to identify the demand regimes favoring each system and the cost differences. Simulations are useful as a guidance to choose systems configurations.

For high demand levels, the total cost per passenger for fixed route and checkpoint systems is very close. In fact their optimal configurations are so alike, and the occurrence of route deviations is so rare, that fixed route systems should be preferred, as they can be operated on a schedule and require less dispatching effort. As the demand level decreases, demand responsive systems become relatively more attractive than fixed route systems, and checkpoint systems might possibly become cost-effective. However, by the time demand responsive systems are significantly better than fixed route systems, door-to-door service can be provided at an even lower cost. This appears to limit the situations where checkpoint dial-a-ride systems can be applied efficiently to a small window of demand levels (and special situations) where they only narrowly outperform other systems.<sup>7</sup>

Casey R., Porter C., Buffkin T., Hussey L. (2000) report on the Cape Cod Regional Transit Authority (CCRTA) Advanced Public Transportation System (APTS) project. It is an application of ITS to fixed-route and paratransit operations in rural transit setting. The purpose of the project is to apply ITS technology that will improve intermodal transportation services. While the paratransit/dial-a-ride system serves residents only, because of the summer tourist pattern of the country, the fixed-route services experience significant seasonal changes in demand. The tools to be implemented are the same as usual in such systems. The integration and compatibility issues are also discussed, particularly as regards the payment technology. This plan describes several measures to be undertaken as part of the whole ITS program related to 4 main goal areas (safety, mobility, efficiency and productivity). Moreover, it provides some practical parameters and other estimators corresponding to those measures and reflecting more directly measurable benefits. The evaluation relies upon analysis of transit agency operational data, including archived historical data, customer survey data as well as financial data on system costs. In addition, interviews with transit agency staff have served to identify impacts that cannot be quantified. The benefits of this evaluation are expected to demonstrate the viability of APTS technologies in addition to the knowledge of the project's effectiveness itself. It might help other operators in assessing the full range of benefits and costs of APTS technologies, both for themselves and their customers. This paper also considers the agency's staff point of view, ease of work and time productivity in addition to the more commonly addressed customers satisfaction topic. Furthermore, it advises to evaluate some additional issues such as the functionality of the APTS system (reliability, accuracy, ease of use), the institutional arrangements or considerations leading to specific problems or successes, and the local environmental factors (geography, travel patterns...) that might have some impacts.

Crum M., Kihl M., Shinn D., (1996) assess the benefits to a small rural DRT operation of investing in smart technologies, assess the relative benefits of investments in AVL or dynamic scheduling,

---

<sup>7</sup> Note that for dynamic problems when the activity of stops changes as the tour is being constructed (requests in real-time), an optimal routing strategy (if one could be found!) should be less efficient than the static TSP case.



assess the potential increases in benefits of a system that combines real-time vehicle tracking with dynamic scheduling, and assess the level of benefits resulting from coordinating acquisition of technology across a consortium of small transit systems (also coordination of operations).

The two major questions to be addressed are: is new technology (AVL and dynamic scheduling here) cost effective? and do the benefits in increased effectiveness and efficiency justify the investment in technology? To begin this paper describes the needs of small DRTS and the potential response offered by advanced transit technologies. The potential benefits of using technology are, for instance, the reduced dispatching and scheduling cost, the decreased vehicle operating cost, the increased revenue, but there are still some intangible or hard-to-measure benefits (non monetary elements). It also describes AVL and dynamic scheduling systems to be used in the context of a DRTS.

In order to assess all the benefits and costs, it uses a BCA model (including two main parts: the projected operating efficiency gains and the projected ridership and net revenue gains), and studies different cases (purchase versus leasing, AVL versus dynamic scheduling with different levels of increase). The model is explained step by step and the payback period of investing in new technology is also considered.

Practically, some primary considerations in choosing scheduling software are addressed (such as service requirements and expectations, area and population density, fleet size, data collection requirements, funding level available). The paper also provides some examples of performance standards (pax/hr, pax/mile, ride time, wait time, deviation time, complaints), productivity measures (cost/revenue-hr or mile, cost/vehicle-hr or mile, cost/pax, pax/revenue-hr or mile, maintenance cost/veh-hr or veh-mile, avg # of veh scheduled/hr), and useful operating data (total operating cost, veh-hr and miles, revenue-hr and miles, cancellations, fuel consumption, maintenance cost, booked trips, # riders). It provides a possible solution to efficiently link AVL and dynamic scheduling systems.

Regan A., H. Mahmassini, P. Jaillet (1998) introduce the problem of dynamic fleet management for truckload carrier fleet operations, and describe the principal components of such operations (the dispatch center, the fleet of vehicles, the communication system, and the transportation system) as well as a simulation framework for the evaluation of dynamic fleet management systems. The application of the simulated framework to the investigation of the performance of a family of real-time fleet operational strategies, which include load acceptance, assignment and reassignment strategies, is also described. The simulation framework presented is an example of a first-generation tool for the evaluation of dynamic fleet management systems (example of performance measures such as operating efficiency measures and quality of service measures are given). The most challenging aspect of those simulations is to define the operational strategies, each one consisting of a load acceptance strategy coupled with an assignment strategy. Two modification strategies are also considered: the en route diversion and the real-time load swapping. Assignment strategies can be more or less flexible: first called, first served; nearest point; classical bipartite with an accumulation of customers into a pool till a specific trigger point (a flag-time for instance) when customers are then assigned to vehicles; en route diversion with or without reassignment that is with or without inter-vehicles and intra-vehicles exchanges. The load acceptance strategies are of two kinds: feasibility based versus profit based.

Different scenarios are simulated to examine the performance of operational strategies under varying demand levels, for vehicle fleets of varying sizes, and with binding or loose capacity constraints. Selected experimental results (from comparisons of different scenarios) are highlighted. Those are intended to illustrate some of the issues encountered in real-time fleet management and the role of the simulation modeling environment in investigating them.

Regan A., H. Mahmassini, P. Jaillet (1996) examine the application of intelligent transportation system technologies to freight mobility. They note it requires dynamic decision-making techniques for commercial fleet operations, using real-time information. Recognizing the productivity-enhancing operational changes possible using real-time information about vehicle locations and demands (several technologies available are described) coupled with constant communication between dispatchers and drivers, a general carrier fleet management system is described in this paper. The purpose of the study is to identify and test ways in which operations should change to take full advantage of real-time information on vehicle locations and demands, since the ability to make decisions dynamically is a key to provide a reliable and efficient responsive service to time-sensitive demands at a reasonable cost.

The system features dynamic dispatching, load acceptance, and pricing strategies. A simulation framework is developed to evaluate the performance of alternative load acceptance (rules which more or less restrictive depending on the level of service aimed) and assignment strategies (two different scenarios: loads held in a large pool of accepted demands until their assignment when a flag-time is reached versus accepted demands assigned directly to a particular vehicle's queue) using real-time information.

Real-time decision making for fleet operations involves balancing a complicated set of often conflicting objectives. The simulation framework provides a means for exploring the trade-offs between these objectives. Two strategies are presented:

- The diversion strategy: diversion to an alternative pick up point of a vehicle which is en route to a specific location, thereby inducing a re-sequencing or reassignment of the original load.
- The repositioning strategy: instead of movements routed over least-distance paths, vehicles could be routed through regions of high demand in hope that demands for service will materialize along the way (think about a least-cost path algorithm in which the costs on links reflect the likelihood that a revenue-generating demand for service will be generated in the near future).

The paper focuses on the diversion strategy consisting of diverting a vehicle en route to a specific location to make a pick up of a more time-sensitive load or a load that when sequenced first will improve the efficiency of the vehicle's travel route. The question is whether this strategy, which requires additional flexibility, will increase the efficiency and/or profitability of the system?

Simulations allowing comparisons (between the diversion scenario and a base case in which demands are served in order they arrive or under different load acceptance rules...) show that the diversion strategy performs well generally. Results suggest that reductions in cost and improvements in service quality should result from the use of dynamic dispatching (assignment) strategies in addition to traditional planning tools.

In a 1995 paper, Regan A., H. Mahmassini, P. Jaillet, investigate the advances in communication, AVL, and GIS technologies have made available several types of real-time information with

benefits for commercial vehicle operations. Continuous updates on vehicle locations and demands create considerable potential for developing automated, real-time dispatching systems.

This paper describes the technologies that are available for use in commercial vehicle operations, explores the potential benefits of a diversion strategy in response to real-time information under idealized conditions as well as the factors affecting those benefits, identifies and designs the strategies that worth consideration, and presents selected results derived from the simulations.

These results illustrate potential savings (reduced travel distances and thus improved efficiency) from simple diversion strategies under real-time information (compared to a avaricious strategy where vehicles are sent to the nearest point without further considerations), as a result, demonstrate the potential power of reacting to even small amounts of real-time information on the state of the system, and highlight the need for methodological development under more realistic scenario to support improved truckload carrier operations decisions.

## Methodology

We begin with a fixed exogenous demand which is homogenous.<sup>8</sup> The objective function may be either to maximize customer welfare or satisfaction or to minimize costs. The variable and parameter definitions are:

- a given area  $A$  (sq miles)  $\geq 0$ .
- $N$  (integer) customers per day spread over  $A$  randomly (uniform distribution for example).  $N \geq n \geq 0$ . Each customer has specific attributes, so we also have:

- $N$  coordinates 'in' at origin  $(X_{in}, Y_{in}, t_{in})$
- $N$  coordinates 'out' at destination  $(X_{out}, Y_{out}, t_{out})$
- $N$  service times requested  $t_n$  (discrete time)

Note that  $t_{out}$  is not requested by users, it is an observed time and, ideally,  $t_{in} = t_n$ .

- for each customer  $n$ , variables 'in' and variables 'out' are assigned to the same vehicle  $k$ . Furthermore,  $n$  is dependent on  $k$  through variable  $X_{nk}$  defined by:

$$X_{nk} = 1 \text{ if } n \text{ is in } k$$

$$X_{nk} = 0 \text{ otherwise}$$

- $K$  (integer) vehicles covering  $A$ .  $K \geq k \geq 0$ . Homogenous fleet of capacity  $C=15$  seats (integer).  $K$  and  $C$  are set but can be changed depending on availability, objectives, operations as well as other factors.<sup>9</sup>

- uniform speeds set =20 mph. We can introduce a variance later to reflect the traffic conditions variability (see congestion issues), which should affect the service quality.

- TW size is the same for each customer, set at 30 minutes.

- Service duration per day  $L$  depends on operators' policy and drivers hours.

---

<sup>8</sup> Later other patterns will be introduced such as: a stochastic demand, a dynamic (in real-time) system, a larger area, a non-homogenous demand and fleet of vehicle.

<sup>9</sup> Small size=taxi versus large size=transit.

- Euclidean distance (miles)  $d_n \geq 0$  with  $d_n = \sqrt{(X_{out} - X_{in})^2 + (Y_{out} - Y_{in})^2}$
- Number of stops per vehicle (integer)  $S_k \geq 0$ . We have two stops per customer (pick-up & drop-off).
- Total distance traveled per vehicle  $k$  (miles)  $D_k = \sum_{n \in N} X_{nk} * d_n \geq 0$ .
- Total distance traveled (miles)  $TD = \sum_{k \in K} D_k$ .
- Time per stop  $t_s \geq 0$ . We assume it is the same for each customer since the demand is homogenous and  $t_s = 5$  min.
- Base time  $tb_n = d_n / s \geq 0$ . This is the time needed for the direct trip of a customer going from  $O$  to  $D$ .
- Deviation time  $td_n = |t_{in} - t_n| \geq 0$  and  $td_n \leq 30$  min (the vehicle can be late or in advance).
- Extra in-vehicle time  $tx_n = (t_{out} - t_{in})_n - tb_n \geq 0$  with  $(t_{out} - t_{in})_n = tr_n$ , riding time.

Additional assumptions are:

- Non-FIFO process.
- No difference between peak and off peak hours.
- Times are expressed into the same unit.
- Assume a sufficient demand for continuous tour of vehicles (no idle vehicle).
- A unique fee per customer (no time or distance related).
- We have a list of times  $t_n$  related to customers  $n$  (sorted by increasing order).
- Each customer  $n$  is characterized by a time  $t_n$  and coordinates  $(X_{in}, Y_{in})_n$  and  $(X_{out}, Y_{out})_n$ .
- For the assignment of customers to vehicles, we use an insertion algorithm. This will produce clusters of customers (2 clusters for O&D).

## ALGORITHM DESIGN

The algorithm follows the sequence:

- 1) Select a customer  $n$  that is, a time  $t_n$  from the list
- 2) For each vehicle  $k$  available (from  $k=0$  to the last vehicle added to the list):

Find all the feasible places where  $n$  can be inserted into the tour/route of the vehicle  $k$ ,

subject to:  $t_n \leq t_{n+1}$  (time relevancy)

$\sum_{n \in N} X_{nk} \leq C_k \forall k$  (capacity constraint); with  $C_k = C \forall k$  for the moment.

$td_n \leq 30$  min (user satisfaction)

$tx_n \leq tb_n$  (that is  $tb_n \leq tr_n \leq 2tb_n$ ) (user satisfaction)<sup>10</sup>

When feasible places are identified for the customer  $n$  within all vehicles  $k$  tours, find the optimal insertion location by minimizing the insertion cost as follow:

---

<sup>10</sup> This condition has turned out to be the most challenging.

Min  $TT = \sum_{k \in K} \sum_{n \in N} \{X_{nk} * [tb_n + t_s + tx_n + td_n]\}$  (the total travel time for all vehicles in service)

s.t.  $t_n \leq t_{n+1}$

$\sum_{n \in N} X_{nk} \leq C \quad \forall k$

$td_n \leq 30 \text{ min}$

$tx_n \leq tb_n$

$\sum_{n \in N} [tb_n + t_s + tx_n + td_n] \leq X_{nk} * L \quad \forall k$  (each trip of each customer of a vehicle should be smaller than the service day duration of this vehicle)<sup>11</sup>

We obtain  $TT^*$  and can deduce  $k^*$ .

If not feasible, immediately take vehicle  $k+1$  and begin the step again.

If not possible, hire a new vehicle with respect to  $k \leq K$  and redo the step or reject this customer/demand (rejection when at capacity or when a customer is considered too costly).

3) Assign  $n$  to  $k^*$  and update information:

- Number of stops per vehicle  $S_k$ .
- Total distance traveled per vehicle  $k$   $D_k$ .
- Total distance traveled  $TD$ .
- Total travel time  $TT$  (should be minimum from the optimization).

If the list is not empty, go to step 1) otherwise stop the algorithm.

## SIMULATIONS

Four groups of simulations were carried out:

1. Vehicles have a capacity of 8 passengers.

The area is a square of 6\*6 miles.

People are picked up in the order of their requested times (increasing  $t_n$ ). And, they are delivered with respect to an optimal TSP (minimum travel time/distance).

2. Vehicles have the same capacity.

The area is the same.

People are dropped off and picked up in the same order, that is the one of their requested times (FIFO process).

3. The capacity is decreased to 5 passengers per vehicle.

The area is still the same.

---

<sup>11</sup> Note the four first constraints have already been checked for the insertion feasibility.

The FIFO process has been kept.

4. Vehicles have a capacity of 5 passengers.

The area has been increased to a square of 12\*12 miles.

The collection and distribution phases are a FIFO process.

There are 150 customers uniformly spread over the area with requested times ( $t_n$ ) uniformly spread over the service horizon (from 8am to 8pm).

## RESULTS:

### Group 1)

In the intermediate solution, 6 customers are rejected because they cannot be served during the service horizon (requested time  $t_n$  + base travel time  $t_{bn}$  ends after 8pm). The simulation is run in order to be able to serve all the remaining customers (=144 persons). This requires 25 tours.

- 4 tours delivering only 1 passenger (= taxi)
- 4 tours delivering 2 passengers
- 1 tour delivering 4 passengers
- remaining tours (=16=64%) are full (= 8 passengers).

The cycles (collection + distribution phases) usually last between 2 and 3 hours for the tours that are full. Several are longer than 3 hours (from 205 min. to 236 min.) or up to 4 hours (287 min.).

The "in-vehicle" times, which reflect the level of service provided, are quite high. The maximum "in-vehicle" times among the different tours vary from 83 min. to an upper bound of 265 min. with the majority greater than 2 hours. Nevertheless, the average "in-vehicle" times are much lower (from a little less than 1 hour to no more than 100 min.). This underlines a large magnitude among those "in-vehicle" times that is among the service provided to customers (some spend hours traveling whereas others are delivered quite quickly). As a result, the FIFO process has been considered rather than the optimal TSP, in order to provide a more equal service among customers and maybe even a better level of service as a whole.

8 vans are needed to service those 25 tours.

No more than 5 tours in a row can be done by a single vehicle.

The total distance traveled by each van is given in the following table in miles:

van	1	2	3	4	5	6	7	8
distance	120.4	116.7	128.6	90.6	78.1	84.9	45	7.5

The total distance traveled to serve the 144 customers over the area is 671.9 miles.

Note that a van begins the first tour at the depot (hypothetically located at the center of the square:  $x=0,y=0$ ) and goes back at the same depot at the end of its service. Those distances from and to the depot are taken into account in the sum. The simulation is based on the optimization of the TSP for the distribution phase.

Group 2)

As before and for the same reason, 6 customers are rejected.

25 tours are still needed:

- 4 tours delivering only 1 passenger (= taxi)
- 1 tour delivering 2 passengers
- 4 tours delivering 3 passengers
- 1 tour delivering 6 passengers (almost full!)
- the others (=15=60%) are full (=8 passengers).

Logically, the cycles (collection + distribution phases) last a little longer than before (from 15 min. to 55 min longer with an average of about 30 min. longer. Surprisingly, one cycle is hardly shorter). More cycles are longer than 3 hours, and there is still one greater than 4 hours (263 min.).

The "in-vehicle" times are still high but with decreased upper bounds. The maximum "in-vehicle" times among the different tours is varying from 81 min. to an upper bound of 251 min. with a majority still greater than 2 hours. The average "in-vehicle" times are a little higher as before (varying from 68 min. to no more than 125 min.). Customers are treated more equally since the magnitude has been decreased as a result of the FIFO process (most of the passengers encounter a higher "in-vehicle" time but the ones who were dealing with the maximum values have seen their "in-vehicle" times decreased).

9 vans are needed to service those 25 tours. One additional vehicle is required as a result of the non-optimization of the TSP. No more than 4 tours in a row can be done by a single vehicle. Note that one of those vans is doing only 1 tour, which at least delivers 6 customers.

The total distance traveled by each van is given in the following table in miles:

van	1	2	3	4	5	6	7	8	9
distance	127	110.5	118.1	102.1	84.5	91.5	110.1	43.2	38.5

The total distance traveled to serve the 144 customers over the area is 825.76 miles.

This is many more miles than with the optimal TSP (154 miles in addition). But the extra cost for the operator results more from the additional vehicle rather than the extra mileage. On the other hand, the level of service is a little better since nobody spends more than 251 min. in the vans (about 15 min. of improvement). This is the trade-off between supplier cost and customer cost.

Consider that  $C_1$  is the cost of a vehicle carrying 8 passengers. From the two previous trials (similar in every other way), it is possible to evaluate the operator's cost (at least a part of it):

$$\text{Cost group1) = } 8 \cdot C_1 + \alpha \cdot 671.9$$

$$\text{Cost group2) = } 9 \cdot C_1 + \alpha \cdot 825.76$$

with  $\alpha$  accounting for the fuel and other mileage costs. We assume  $\alpha = \$ 0.3$ .

$$\text{Cost group1) = } 8 \cdot C_1 + 201.6$$

$$\text{Cost group2) = } 9 \cdot C_1 + 247.7$$

The extra cost of group2 is equal to  $(C_1+46)$  dollars. The question is whether the 15 min. improvement worth more or less than  $C_1$  (that is the additional van).

### Group 3)

6 customers are rejected.

35 tours are needed:

- 5 tours delivering only 1 passenger (= taxi)
- 3 tours delivering 2 passengers
- 1 tour delivering 3 passengers
- the others (=26=75%) are full (=5 passengers).

Hopefully, the cycles (collection + distribution phases) are shorter than before since the capacity has been decreased. They last usually between 1 and 2 hours, more precisely between 1,5 and 2 hours. Some (about 1/3) are longer than 2 hours, and there is even one greater than 3 hours (190 min.).

The "in-vehicle" times are more acceptable. The maximum "in-vehicle" times among the different tours is varying from 47 min. to an upper bound of 156 min. with a majority between 60 and 80 min. The average "in-vehicle" times are usually close to 1 hour (varying from 43 min. to 85 min.). Interestingly, by decreasing the vehicles' capacity, the level of service has been greatly improved.

8 vans are needed to service those 35 tours. It is thus possible with the same number of smaller vehicles to run more and shorter tours in order to achieve a better level of service by decreasing the "in-vehicle" times of the travelers.<sup>12</sup>

No more than 6 tours in a row can be done by a single vehicle. Note that 2 of those vans are doing only 1 tour (full at least!). Since those tours encounter the highest maximum "in-vehicle" times, the related customers (= 10 persons) might be rejected by the operator. The total distance traveled by each van is given in the following table in miles:

van	1	2	3	4	5	6	7	8
distance	160.8	138.5	133.2	132.8	139.2	57.4	38.9	32.7

The total distance traveled to serve the 144 customers over the area is 833.70 miles. This is approximately the same total distance as before, which is logical since the process is the same as well as the area covered.

---

<sup>12</sup> In the limit this would revert to a pure taxi model.



Consider  $C_2$  is the cost of a smaller vehicle carrying only 5 passengers.

Logically,  $C_1 > C_2$ .

From the two previous trials (similar in everything else than the capacity), it is possible to evaluate the operator's cost (at least a part of it):

$$\text{Cost group2) = } 9 \cdot C_1 + \alpha \cdot 825.76 = 9 \cdot C_1 + 247.73$$

$$\text{Cost group3) = } 8 \cdot C_2 + \alpha \cdot 833.70 = 8 \cdot C_2 + 250.11$$

Clearly the alternative group3 is preferable because of a lower cost for the operator and a better level of service for the customers who spend on average 30 min. less in the vehicle.

#### Group 4)

In this trial, more customers are rejected since the area has been extended and the service standard and inputs are still fixed (indeed, the base travel times  $t_{bn}$  are greater, and thus more people fail to receive service; 139 customers remain to be serve.

34 tours are needed:

- 4 tours delivering only 1 passenger (= taxi)
- 3 tours delivering 2 passengers
- 2 tours delivering 3 passengers
- 2 tours delivering 4 passengers (almost at capacity!)
- the others (=23=67%) are full (=5 passengers).

Logically the cycle lengths have increased since the distances covered are greater. The cycles last between 2 and 4 hours depending on the tours, with some exceeding 4 hours (248 min. or even 288 min.).

The "in-vehicle" times are still acceptable considering the distances traveled since the capacity has been kept equal to 5 passengers. The maximum "in-vehicle" times among the different tours is varying from 86 min. to an upper bound of 172 min. with a majority between 110 and 140 min (that is around 2 hours). The average "in-vehicle" times are composed of trip time limits of somewhere between 1.5 and 2 hours (varying from 66 min. to 147 min.). By doubling the area serviced, the "in-vehicle" times are greatly increased (not fare from doubled!). On the one hand the level of service is improved through better coverage but on the other hand this level of service is diminished because of greater passengers' "in-vehicle" times.

11 vans are needed to service those 34 tours. Even if more vans are needed, it is not necessary to double the fleet when the area of service is doubled. But no more than 4 tours in a row can be completed by a single vehicle.

Note that 1 van is doing only 1 tour, albeit full, and another only 2 tours (also near full). Once again two of those tours requiring additional vehicles encounter the highest maximum "in-vehicle" and average times, showing a relatively bad level of service.

The total distance traveled by each van is given in the following table in miles:

van	1	2	3	4	5	6	7	8	9	10	11
distance	183.9	149.2	184.7	147.6	168.9	158.2	131.6	138.8	103.4	137.5	49.2

The total distance traveled to serve the 139 customers over the area, which is twice larger, is 1553.33 miles (what is not twice as great as the total distance traveled in the previous trial).

#### NEXT STEPS

Once a feasible set of solution algorithms with the variety of conditions we described in the beginning has been developed, we can introduce AVL or some other type of ITS system. The fundamental way in which any ITS implementation will work is to change information access. More complete and timely information serves to provide opportunities to better deploy resources and have customers better utilize their time. It can also reduce the anxiety of not knowing when things are going to happen. Introducing ITS will yield a different solution outcome and the characteristics of the two solution equilibriums provide a measure of the net benefits of such an ITS investment. However, before that can be pursued we need to better understand how we can model the maximization of welfare objective. If this is not done we restrict the benefits of ITS to merely reducing costs and not increasing satisfaction and customer service. In effect we need to be able to value customer service at different levels and changes to those levels.

#### CONCLUSION:

The marginal cost of additional vehicles or changing the process of operating for example, and the marginal benefit from time savings, which is somewhat more difficult to evaluate, depend on the strategy chosen by the operator considering the money available and the level of service desired.

The objective of the operator would be to minimize operating costs while providing a service which adheres to some minimum standard. Practically, the objective is to minimize the cost by minimizing the number of vehicles and the total distance traveled, while maximizing the level of service by delivering customers over a large enough area within a reasonable amount of time, that has an acceptable "in-vehicle" time. Note that with this model and those simulations, it is not question of waiting time (at home or wherever), but of "in-vehicle" time, which is better perceived and accepted by people (the feeling of wasted time is greater when waiting doing nothing rather than being in a van to be driven somewhere).

The four groups of trials show the different trade-offs available to the operator. Depending on the amount of money they are willing to spend in order to provide a specific level of service, they might chose one or another of those alternatives. The point would be to conduct a market survey to understand people's expectations and especially people's value of time in order to model the "best" service, or at least to get closer to this best solution at a lowest cost. In principal what is needed is some sort of stated preference survey that would reveal a willingness to pay for a variety of different service levels.

## References

- Casey R., Porter C., Buffkin T., Hussey L., *Evaluation plan for the Cape Cod Advanced Public Transportation System*, June 2000, 49p, <http://www.itsdocs.fhwa.dot.gov/>
- Crum M., Kihl M., Shinn D., *Linking Real Time and Location in Scheduling Demand Responsive Transit* (final report), by., September 1996, 124p.
- Daganzo, C. An Approximate Analytic Model of Many-to-Many Demand Responsive Transportation Systems, *Transportation Research*, vol. 12, No. 5, 1978, pp. 325-333
- Daganzo, C. , Checkpoint Dial-a-Ride Systems, *Transportation Research*, vol. 18B, No. 4/5, August/October 1984, pp. 315-327.
- Regan A., H. Mahmassani, P. Jaillet, Evaluation of Dynamic Fleet Management Systems *Transportation Research Record*, No. 1645, 1998, pp. 176-184
- Regan A., H. Mahmassani, P. Jaillet, Dynamic Decision Making for Commercial Fleet Operations Using Real-Time Information, *Transportation Research Record*, No. 1537, November 1996, pp. 91-97
- Regan A., H. Mahmassani, P. Jaillet , Improving Efficiency of Commercial Vehicle Operations Using Real-Time Information: Potential Uses and Assignment Strategies, *Transportation Research Record*, No. 1493, July 1995, pp. 188-198.