

# On-Line and Off-Line Routing and Scheduling of Dial-a-Ride Paratransit Vehicles

Liping Fu\*

Department of Civil Engineering, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1

&

Stan Teply

Department of Civil Engineering, 215B Civil/Electrical Building, University of Alberta, Edmonton, Alberta, Canada T6G 2G7

**Abstract:** *This article presents the general concepts, models, and computational techniques applied in a new dial-a-ride vehicle routing and scheduling system. The objective of this system is to improve the responsiveness, reliability, and productivity of dial-a-ride paratransit services. The developed software integrates dial-a-ride routing and scheduling principles and practical experience and explicitly considers travel time variability in urban roadway networks. Such extensive and complex integration has been made possible by improved data acquisition and processing capabilities of computer, telecommunications, and vehicle location technologies. Advanced computational methods applied in the system, such as the artificial neural network technique, which allows heuristic estimation of origin-destination travel times in a dynamic and stochastic fashion, contribute to the processing speed required to respond expeditiously and efficiently to paratransit user requests. A real scheduling problem from the city of Edmonton, Alberta, where the system was tested, is used to illustrate the positive computational experience and the capability of the developed software to handle both off-line and on-line operations.*

\* To whom correspondence should be addressed. E-mail: [lfu@uwaterloo.ca](mailto:lfu@uwaterloo.ca).

## 1 INTRODUCTION

As a public service, paratransit systems have played a special role in providing transportation to a population with reduced mobility, such as the elderly and handicapped who have difficulties in accessing the regular fixed-route public transit system. Due to the door-to-door service characteristics, most paratransit systems rely heavily on subsidization. The latest development in intelligent transportation systems (ITS), such as automated vehicle location (AVL) and digital communications, have the potential to make these systems more cost-effective.<sup>3,6,14</sup> These technologies allow an operation center to constantly track the service vehicles and to effectively communicate with drivers and customers. In addition, continuing advances in computing power have increased data-processing ability so that more complex models can be applied and schedules updated in real time. The efficiency, productivity, and reliability of paratransit service can be improved significantly, provided that increased data acquisition and processing needs can be met and the resulting information used adequately in the management and operations processes. The objective of this article is to present an overview of the concepts, models, and computational technologies applied in a new dial-a-ride vehicle routing and scheduling computer software program that has been developed with special consideration of the increased data availability.

This article first presents an overview of various classes of dial-a-ride problems (DARP) involved in a typical dial-a-ride paratransit system and a conceptualization of an Automated Dial-a-ride Paratransit Operation System (ADPOS). The major components and features of the program developed in this research are then introduced, focusing on a new DARP model that explicitly considers the dynamics and stochastic nature of the origin-destination (O-D) travel times and estimation of the O-D travel time based on an artificial neural networks (ANN) technique. Finally, the implementation issues involved in the program are discussed and illustrated using a real application problem from the Disabled Adult Transportation System (DATS) of the city of Edmonton, Alberta, Canada.

## 2 BACKGROUND

### 2.1 Dial-a-ride vehicle routing and scheduling problems

In a dial-a-ride paratransit system, one of the daily operational problems facing the service operator is to determine the optimal pickup and drop-off routes and schedules for a fleet of vehicles to carry customers between specified origin and destination locations. This problem belongs to the well-known vehicle routing and scheduling problem in operations research (OR) and is specifically referred to as the *dial-a-ride problem* (DARP).<sup>1,13</sup>

There are fundamentally two DARP types involved in the operation process of a dial-a-ride system that need to be treated differently. The first problem is the *static DARP*, which usually needs to be solved at the beginning of every day when all the customer trip requests are known (e.g., booked one day in advance or subscribed for regular service). The objective of the routing and scheduling procedure is to determine the assignment of all customers (or trips) to the available vehicles and their respective routes and schedules. This function can be performed off-line and thus is less time demanding.

The second problem in a dial-a-ride system is called *real-time DARP*, as referred to in this article, in which the objective is to determine the assignment of customers into the existing schedules of vehicles in real time (these trips are also called *demand trips*). The new customers usually phone the service center to request immediate service while some fleet vehicles are already in service and have a set of given operation schedules. Because of the requirement imposed by a real-time operation process, the dispatcher usually has to give the customers their pickup/drop-off times immediately; therefore, the “optimal” insertion schedules must be found in a very short time period (e.g., a few seconds). A fast on-line routing and scheduling algorithm is therefore required to solve this problem.

The DARP can be further classified into various types

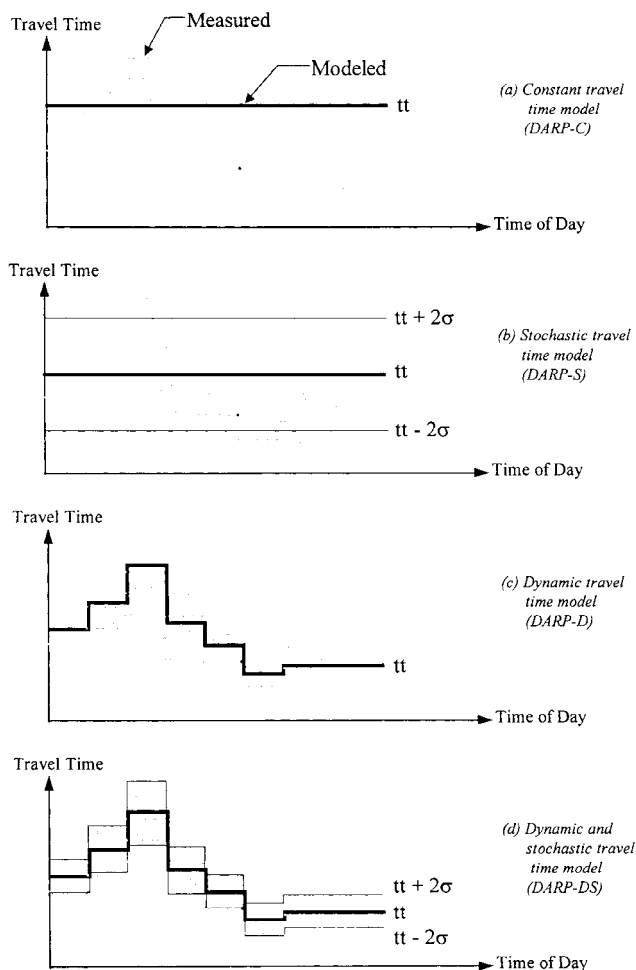


Fig. 1. Schematic illustration of O-D travel time models.

of subproblems based on how the system components and parameters are represented or modeled. This article classifies the DARP into four problems based on how the origin-destination (O-D) travel times are modeled (Figure 1). The first class is defined as dial-a-ride problems with constant O-D times (DARP-C); that is, the travel time between each O-D pair has no variation and is independent of the time of day, as shown in Figure 1a. The second class of DARP, denoted as DARP-S, considers the stochastic variation of the O-D travel times by modeling them as random variables (Figure 1b). The third class of DARP, denoted as DARP-D, assumes that the O-D travel times are dynamic, that is, time-dependent but not stochastic, as shown in Figure 1c. When the travel times are modeled as both dynamic and stochastic, as shown in Figure 1d, the result is the last class of problems—DARP-DS.

The DARP can be modeled differently in terms of objective functions and constraints considered in problem formulation. These objectives and constraints are related to service providers and service clients—customers. Examples of the

objective functions include

*Objectives related to service providers:*

- *Minimize the number of vehicles.* This objective is related to service providers' operating cost. The operating cost is commonly proportional to the number of vehicles used,<sup>13</sup> especially when some rented or contracted vehicles are used in addition to internal vehicle fleet.
- *Minimize the total vehicle travel time.* The travel time is another important factor that influences the service providers' operating cost.

*Objectives related to customers:*

- *Minimize customer service time deviations.* Customers' satisfaction with the service provided is usually measured in terms of service time deviations, that is, the time difference between the actual or scheduled pickup/drop-off times and their most desired pickup/drop-off times.
- *Minimize customers' excess ride times.* In shared-ride systems, customers normally experience some extra in-vehicle time (excess ride time) as compared with their direct ride time (without diversions for other customers).

When more than one objective function is used, a general cost function or disutility function is commonly formed to combine them together.

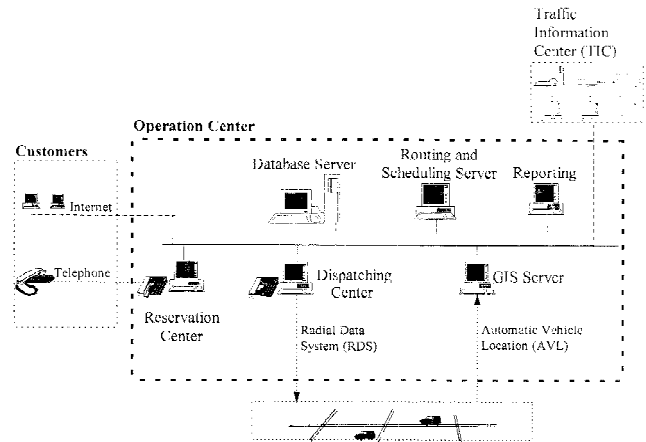
The operational constraints that must be satisfied during the routing and scheduling process include

*Constraints related to service providers:*

- *Limited number of vehicles.* The number of vehicles planned for service is typically a decision made at the higher-level strategic service planning.
- *Vehicle capacity.* Vehicles are limited in capacity by seating types.
- *Vehicle service time periods.* Service vehicles are usually not available all day. Service operators need to control the service plan through the allocation of service time periods, and drivers have to take breaks. Typically, there are multiple service time periods associated with each vehicle.

*Constraints related to customers:*

- *Seating requirements.* Customers usually have different seating requirements. For example, some customers can use regular seats, whereas others have to remain in wheel chairs. During the routing and scheduling process, it must be guaranteed that the number of on-board customers on each vehicle does not exceed the vehicle capacity of each seating type at each stop along the route.
- *Pickup/drop-off time windows.* Customers usually specify either desired pickup or drop-off times. To avoid extremely long waiting times, local service policy usually designates a maximum time deviation from customers' desired pickup/drop-off times. As a result, customers must



**Fig. 2.** The Automated Dial-a-Ride Paratransit Operation System (ADPOS).

be scheduled to be picked up or dropped off at specific time periods—time windows.

- *Maximum ride time.* In order to avoid extremely large excess ride times for each customer in a schedule, a maximum ride time is commonly specified as a routing and scheduling criterion.

## 2.2 Automated dial-a-ride paratransit operation systems (ADPOS)

Dial-a-ride paratransit operation requires a set of operating functions such as trip reservation, vehicle routing and scheduling, vehicle dispatching, and business reporting. With various technologies such as microcomputers, AVL, and digital communication, these operation functions can be fully automated for potentially more cost-effective operations.<sup>2,10,14</sup> This section presents a conceptualized system called *Automated Dial-a-ride Paratransit Operation System* (ADPOS) with an assumption that various state-of-the-art technologies are available, as shown in Figure 2. The focus is on how the system works and what type of information is available for use in the routing and scheduling process.

In this hypothetical ADPOS shown in Figure 2, most components can be fully automated. At the start of the service day, routes and schedules are automatically generated by a computer (scheduling server) and sent to each driver. Most customers request service in advance through the Reservation Center. During the service day, some customers call the Dispatching Center to request immediate service as opposed to prearranged or scheduled service. As soon as the Dispatching Center receives customers' identification numbers or names, their registration information is verified at the database server. A request for inserting these new trips into the existing schedules is then sent to the scheduling server. Within a few seconds, the schedule server determines which vehicles to assign

to the trips based on current vehicle locations (from GIS and AVL) and traffic condition (from TIC), calculates the new vehicle schedules, and informs the Dispatching Center when the vehicles will provide the service to the customers. The Dispatching Center then sends the revised schedules to the drivers through a digital communication system such as Radio Data System (RDS). It should be noted that the diagram shown in Figure 2 represents only the functional configuration of the ADPOS and should not be interpreted as a definite software or hardware setting. In fact, many of these components can be combined and performed by a single computer. The following subsections provide a more detailed discussion of each component.

*Reservation Center.* The Reservation Center provides the connection between the Operation Center and customers. This center takes responsibility for recording the customer reservations, including pickup and drop-off locations, desired time window, number of passengers to be delivered (or picked up), and their special requirements; it is also responsible for updating customer information in the case of service cancellation and change of service time. This procedure is usually conducted by reservation clerks who receive the information over the telephone and then manually enter it into the computer. After the data are analyzed, the clerks inform the customers whether or not the service request can be accommodated. As an option, the reservation also can be completed by an automated telephone system connected to a computer through a modem. Customers can dial in to make their own trip reservations.

Internet technology also may be applied effectively. In this case, a WWW site is set up for customers to access using a browser such as Netscape Navigator or Microsoft Explorer. With a form-based registration system, the customer can register service and reserve trips on-line without the help of service clerks. The front forms for registration and reservation are dynamically linked to the database on the Database Server.

*Traffic Information Center (TIC).* The TIC is a major component of the ITS. Its function includes managing the traffic data from a variety of data sources, such as loop detectors, probe vehicles, computer simulation models, etc., and then distributing the processed information to different applications, such as advanced traveler information systems (ATIS) and ADPOS. The data communication between TIC and ADPOS has several options. The first option is a periodic transfer through portable storage media such as a zip driver and regular floppy diskettes. It is simple and inexpensive but can only provide historical data. It is therefore suitable for transferring a large amount of data. The second option is through a batch file, in which data are generated in the TIC and then transmitted to ADPOS for storage through a computer network. A

more complex alternative is the on-line link through intranet or Internet, that is, the ADPOS directly accesses the data in the TIC. This method provides an opportunity for ADPOS to obtain and update real-time network information without much delay. The major constraint is the limitation of transfer speed during peak information demand times.

*Geographic Information System (GIS) server.* The GIS server contains both explicit geographic references such as latitude and longitude coordinates and implicit references such as street networks, addresses, and postal codes. Geographic relationships can be established between these two types of references, which allow identification of the coordinates of a location based on its address. The coordinates of locations form the basic inputs for calculating travel times between locations that are used in the routing and scheduling process.

*Automated vehicle location (AVL).* A number of alternatives are available to locate and track vehicles. The most basic method is two-way communication (e.g., by cellular phone), with the driver identifying his or her location to the system operators. The second method is based on communication between vehicles and roadside facilities (sign posts). These facilities provide the times when the vehicles pass known locations. Another method is called the *global positioning system (GPS)*, which uses satellites to locate and track the vehicles. Information from AVL can be verified or processed by the GIS server with a real-time display of vehicle locations on the computer screen.

*Database server.* All the data related to the dial-a-ride service operation are stored and managed by a central database management system (DBMS)—the database server. All the information collected by the Reservation Center, TIC, and AVL are input and stored in a central database. These data are managed by a DBMS system such as Oracle or Microsoft SQL server or access.

*Routing and scheduling server.* This server is responsible for providing optimal routes and schedules for the available vehicles. This is the core function of the system. The program developed to perform this function is discussed in more detail in the following sections.

*Dispatching Center.* The dispatchers keep monitoring the system changes for incidents, service cancellations, and new trip requests for immediate service. These changes may justify modification of vehicle schedules, such as

- *Regroup.* New system situations may result in assigning different customers to different vehicles compared with those from the preceding dispatching run.
- *Resequence.* Change the sequence to visit each customer.

- *Add a new vehicle.* When new demands cannot be accommodated or cannot be served efficiently, it may be necessary to dispatch a new vehicle.
- *Redirection.* This operation focuses on the one-to-one routing in a road network.

Once the new operation strategy is verified, the schedules are sent to the drivers using a cellular phone or RDS.

The system conceptualization shown in Figure 2 focuses on the functionality of the system as a whole and the flow of data between system components. It should be noted that such a system could take many configurations depending on the technologies used for collecting and communicating data and human interfaces. For example, the traffic network information could be real-time or historical data; the vehicle location could be identified by AVL, estimated by communication with the driver, or estimated using vehicle routes and schedules stored in the computer; and the system may use micro- or main-frame computers that could be linked locally through local-area networks (LANs) or globally (Internet).

### 3 SYSTEM FUNDAMENTALS

#### 3.1 Motivation for the research

Since the early 1970s, a number of computer algorithms and programs have been developed to solve the DARP and provide assistance to the complicated scheduling and dispatching process of dial-a-ride paratransit systems.<sup>10,14–16</sup> However, one of the major disadvantages of these tools is that they are based exclusively on the assumption of constant O-D travel times and therefore essentially deal with the DARP-C class of problems described in Section 2.1. In an urban traffic environment, however, O-D travel times may be highly time dependent (dynamic) and uncertain (stochastic) because of the inherent fluctuation in travel demands, unpredictable occurrences of traffic incidents, and changes in weather conditions.<sup>11</sup> For example, a travel time study<sup>12</sup> showed that the travel times in the city of Edmonton (with light traffic) had a wide spread of variability, with coefficients of variation ranging from 0.06 to 0.30.

It can be expected that in situations of high uncertainty the service vehicles may not be able to follow the schedules computed on the basis of the traditional models, and thus reliable service may not be guaranteed. For example, based on the assumption of deterministic O-D travel times, it would be feasible to schedule a vehicle to deliver a customer at his or her destination at his or her most desired drop-off time. Nevertheless, the actual drop-off time will not be exactly the desired or scheduled drop-off time because of the randomness of vehicle travel times. The drawback associated with the use of static O-D travel times is more straightforward in the sense that it may result in erroneous and inefficient schedules.

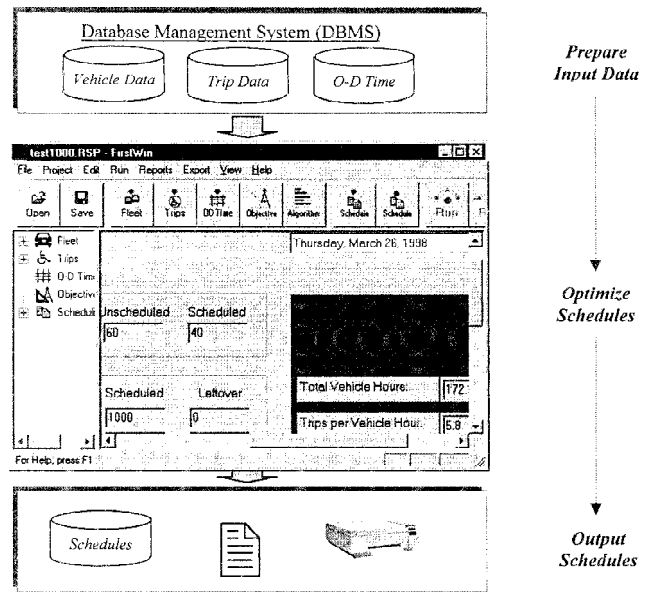


Fig. 3. Structure of the FirstWin routing and scheduling system.

By using static and deterministic models, traditional routing and scheduling tools are not able to take advantage of the large amount of data available due to new technologies. As a result, these tools may not be used satisfactorily in advanced operation systems such as ADPOS. The software system FirstWin developed during this research is designed to overcome the problem. The following subsection presents a detailed discussion on the structure, routing and scheduling models and algorithms, and various O-D travel time estimation methods incorporated into the software.

#### 3.2 Software structure

As a decision support tool, the core function of the software is to find the optimal pickup and drop-off routes and schedules for a fleet of vehicles required to transport customers from various origins to various destinations in a service area. Figure 3 schematically illustrates the operation concept of the system.

The data needed by the program fall into two categories. The first category includes vehicle-related data such as capacity, service time period, and depot location and customer trip-related data such as origin and destination locations (geographic coordinates or zone number), desired pickup or drop-off time and seating requirements, and O-D travel times. These data are prepared externally as database tables using any DBMS that supports the Microsoft Open Database Connection (ODBC) standard. The second category includes scheduling-related parameters such as maximum allowable ride time, maximum service-time deviation, and weights to different optimization objectives, which can be entered and

modified inside the program. The user can import the first category of data from the external database through an ODBC connection and set the second category of data through a user interface. After all the required data are set, several optimization routines can be invoked to assign customer trips to each vehicle. The program will create screen display and text file outputs of schedule summaries and detailed schedules that can be viewed and printed. In addition, the program allows the user to export schedules to a database that can be used later either by itself or with other programs (to generate vehicle run sheets, for example).<sup>4</sup>

### 3.3 DARP models

As discussed in Section 2.1, there are a variety of DARPs that need to be solved in the typical dial-a-ride operation process. FirstWin was developed initially specifically for solving the static DARP and later extended to handle the real-time DARP. A real-time DARP can be viewed as a special case of a static DARP except that the vehicle locations need to be set as the current locations.

While FirstWin was designed to deal with all subproblems based on the O-D travel time models discussed in Section 2.1, the following discussion focuses on the most general and realistic model, that is, DARP-DS.

Due to the involvement of multiple variable objectives and the randomness of system status (i.e., random O-D travel time and thus random arrival time and ride time), the utility concept is used to combine the objectives of the system operator and the customers and to resolve the randomness of the operation measures.<sup>3,9</sup> Specifically, a disutility function, which represents the relationship between the degree of dissatisfaction of the service operator or customers associated with the routing and scheduling results, is used to represent the objectives. Therefore, the general objective function of the routing and scheduling procedure can be represented by the total disutilities of the service operator and customers.

FirstWin considers all the operation constraints discussed in Section 2.1 except that the time-related constraints are expressed as probabilistic conditions because of the randomness of travel times.

The following subsection uses the ride time as an example to illustrate how the objectives and constraints are modeled when the travel time and arrival time are random. A full description of this modeling practice can be found in Fu.<sup>3</sup>

**3.3.1 Objectives and constraints related to customers' ride time.** A chance constraint is introduced to represent the maximum ride time condition under stochastic travel time. The constraint involves the following probability function:

$$\text{Prob}(\tilde{t}_i \leq L) \geq \beta \tag{1}$$

where  $\tilde{t}_i$  is a random variable representing the scheduled ride time for customer  $i$ ,  $L$  is a systemwide parameter represent-

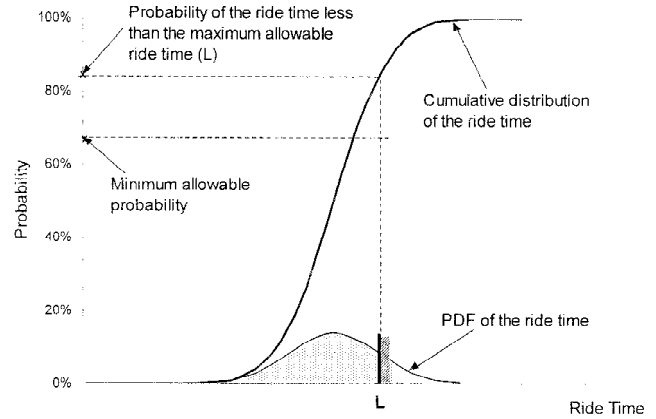


Fig. 4. Customer's ride time condition.

ing the maximum allowable on-board time, and  $\beta$  is an externally set criterion representing the minimum allowable probability. For example, it may be decided that the constraint must be satisfied with a probability of 95%. This constraint is schematically illustrated in Figure 4. The curve represents the probability density function (PDF) of customer's ride time. The shaded area under the curve is the same probability as shown in Equation (1). Equation (1) can be rewritten as

$$\int_0^L f_{\tilde{t}_i}(x) dx \geq \beta \tag{2}$$

where  $f_{\tilde{t}_i}(x)$  = PDF of customer  $i$ 's ride time.

The validation of Equation (2) requires that the PDF or cumulative distribution function of the ride time ( $\tilde{t}_i$ ) for each trip be known in advance. To simplify the problem, it was assumed that the O-D travel time for each O-D pair is normally distributed with mean and standard deviation as known functions of departure time at the origin location. This implies that  $\tilde{t}_i$  also would be normally distributed with its mean and standard deviation obtainable based on O-D travel time data and departure time of the route. Because the cumulative distribution function of a normal variable can be effectively approximated by a simple function, the computing time required for checking Equation (1) can be significantly reduced.<sup>3</sup> It should be noted that the models and algorithms in the program are not limited to the normal distribution assumption.

The maximum ride time criterion provides an upper bound of the service requirement. A scheduling process also aims to minimize customers' ride time. Therefore, a disutility component representing this scheduling objective needs to be included in the general disutility function.<sup>1,8</sup> This disutility component, noted as  $DU_i^r$ , is defined as a quadratic function of the relative deviation of the mean ride time:

$$DU_i^r = a_3 y_i^2 \tag{3}$$

where  $a_3$  = an externally set constant representing the weight allocated to the ride diversion in the general objective function

$y_i$  = the relative deviation of the mean ride time for customer  $i$ , defined as

$$y_i = \frac{E[\tilde{t}_i] - E[\bar{t}_i]}{E[\bar{t}_i]} \quad (4)$$

where  $\bar{t}_i$  = the direct ride time, with expected direct ride time noted as  $E[\bar{t}_i]$

$E[\tilde{t}_i]$  = the expected scheduled ride time for customer  $i$

The reason for using the relative ride time diversion function shown in Equation (4) is that a customer's toleration of the ride time deviation is also related to the direct ride time of his or her trip. Commonly, the longer the direct ride time of customers' trips, the longer is the ride time deviation that he or she can tolerate.

### 3.4 Algorithm

The preceding subsection discussed how the objectives can be modeled using a utility concept and chance constraints. The DARP can then be formulated to minimize the total disutility of service providers and customers subject to a series of constraints such as that shown in Equation (1). A mathematical formulation can be found in Fu.<sup>3</sup> Due to its inherent intractability, only heuristic algorithms are currently feasible to solve practical DARPs. One of the widely used heuristic routing and scheduling procedures is called an *insertion algorithm*, which processes customer trips sequentially, attempting to insert one customer trip at a time into the work schedule of some vehicle while considering all insertion opportunities provided by the available vehicles.<sup>8</sup> This algorithm has been extended to solve the dial-a-ride problem with the new objective functions and constraints discussed earlier. The insertion algorithm has the following iterative solution steps:

- Step 1.* Determine the time windows for each customer.
- Step 2.* Select a customer  $i$  from the customer list.
- Step 3.* For each vehicle  $k$  from the fleet:
  - (a) Find all feasible ways in which customer  $i$  can be inserted into the partial schedule of vehicle  $k$ . If it is not feasible to insert customer  $i$  into vehicle  $k$ , examine the next vehicle  $k + 1$  and restart step 3.
  - (b) Find the insertion of customer  $i$  into vehicle  $k$  that results in a minimum insertion cost.
- Step 4.* If it is not feasible to assign customer  $i$  to any vehicle, then either hire a new vehicle to serve this customer or "reject" this customer. Otherwise, assign customer  $i$  to vehicle  $k^*$  for which the insertion cost is minimal among all the vehicles, and update vehicle schedules.

Three processes are central to the insertion algorithm: travel time updating, a feasibility test to insert a customer trip into an existing schedule, and an optimization process to find the best feasible insertion and best schedule.

*Travel time updating.* During the insertion process, the vehicle schedules are subject to changes because of insertion of new trips, which will in turn change the departure time at some stops. When the O-D travel times are modeled as dynamic, the travel time between stops and the arrival time at each stop must be updated during the insertion process.

*Feasibility test.* The feasibility test is to ensure that the service quality constraints for both the newly inserted customer and all other customers already on that vehicle are not violated. The following tests must be passed for the feasibility check:

- Test 1.* Test the vehicle's load (on-board passengers) at each stop. The vehicle's load always must be less than or equal to the vehicle's capacity.
- Test 2.* Test the satisfaction of the inserted customer and the customers already on the vehicle. This test includes the customers' time windows and the customers' ride times.

*Optimization.* The objective of the optimization step is to minimize the total additional disutility due to inserting a customer into a vehicle's schedule. The additional disutility is the difference between the total disutility before and after inserting a customer. The disutility function is primarily related to the total travel time for each vehicle, the arrival time at each stop, and each customer's ride time. Once a vehicle's pickup and drop-off sequence is determined (e.g., after a customer is inserted into an existing partial route), an optimal starting time for this vehicle can be obtained by minimizing the total disutility of the schedule based on the O-D travel time information. Subsequently, the total travel time, the arrival time at each location, and each customer's ride time can be determined to arrive at the total disutility.

The preceding algorithm also was modified to schedule other types of trips, including trips that are preassigned to specific vehicles by schedulers and group trips that are arranged in advance to be serviced by the same vehicle.

### 3.5 O-D travel time estimation methods

FirstWin allows the O-D travel time to be modeled as constant, dynamic, and/or stochastic, as discussed in Section 2.1. Corresponding to each model, there must be a method available to provide the estimation of the related O-D travel time information. In FirstWin, the following three types of O-D travel time estimation methods are included:

- Zone-based O-D travel time estimation method
- Distance-based O-D travel time estimation method

- Artificial neural network–based O-D travel time estimation method

Each of these individual methods is described below.

**3.5.1 Zone-based O-D travel time estimation.** The service area is first divided into individual zones, and the travel time between each pair of zones is determined in advance and stored in a database table. The travel time between two locations is assumed to be equal to the travel time from the zone where the origin is located to the zone where the destination is located. For convenience, the traffic analysis zone (TAZ) system used in transportation forecasting models may be adopted (as, for instance, in Edmonton, where 515 zones are used).

When the O-D travel time is modeled as a dynamic and stochastic variable represented by its mean and standard deviation as functions of time of day, a number of O-D time matrices for different time periods are required. For computational efficiency, the whole O-D travel time matrix is read in from the database and kept in random access memory (RAM) before FirstWin executes the scheduling algorithms. A large amount of RAM is required for the program to run efficiently. If a system of  $N$  zones is used, the required RAM to keep a single O-D matrix is  $2N^2$  bytes. For the Edmonton case, about 0.5 MB ( $2 * 515^2$  bytes) is required to store a single O-D time matrix.

**3.5.2 Distance-based O-D travel time estimation.** In this method, O-D travel times are approximated by rectangular distance ( $d_1$ ) and Euclidean distance ( $d_2$ ) from the origin location to the destination location. These two distances are determined from the coordinates of the locations as follows:

$$d_1 = |x_o - x_d| + |y_o - y_d| \tag{5}$$

$$d_2 = \sqrt{(x_o - x_d)^2 + (y_o - y_d)^2} \tag{6}$$

where  $(x_o, y_o)$  and  $(x_d, y_d)$  are, respectively, the coordinates of the origin location and destination location.

The relationship between the O-D travel time and these two distances can be established through a regression function:

$$t_{od} = ad_1 + bd_2 \tag{7}$$

where  $t_{od}$  = the O-D travel time from origin ( $o$ ) to destination ( $d$ ), in seconds

$d_1, d_2$  = the rectangular distance and the Euclidean distance, in kilometers

$a, b$  = regression parameters

This regression can be obtained easily using spreadsheet software, such as Microsoft Excel or Lotus 1-2-3, if sample O-D travel time data and their related location coordinates are available. If dynamic and stochastic O-D travel times are used, such regressions have to be established for both mean and standard deviation for each time-of-day period.

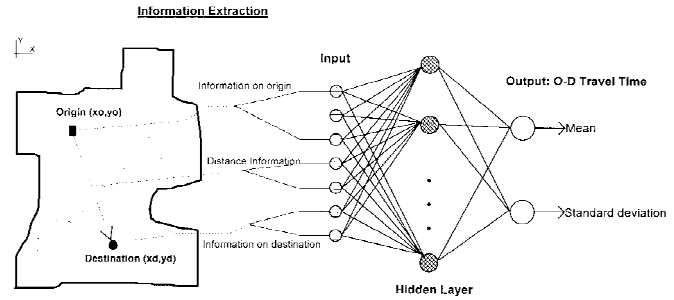


Fig. 5. ANN-based O-D travel time estimation.

**3.5.3 Artificial neural network (ANN)–based O-D travel time estimation.** FirstWin has included a unique method called *artificial neural network (ANN)–based O-D travel time estimation method*. This subsection provides a general overview on the ANN model and describes how it is integrated into the program.<sup>3,5</sup>

The ANN used in FirstWin is known as a backpropagation neural network that can be trained to provide a desired mapping between the O-D travel time (mean and standard deviation) and various input factors such as the coordinates of the origin and destination locations and the departure time at the origin, as shown in Figure 5. Because the ANN model is used as a subcomponent in the routing and scheduling process, a direct code-level integration is necessary for computational efficiency. For this reason, a Windows application program called NeuralBase was developed as a coupled program with FirstWin. NeuralBase provides an interface for training and testing O-D travel time estimation models, whereas FirstWin can use the trained models to estimate O-D travel time during the routing and scheduling process.

The integration between NeuralBase and FirstWin is illustrated schematically in Figure 6. The basic procedure is as follows: A set of neural networks is trained and tested in NeuralBase with data from various sources, such as daily operational records and survey and GIS systems. The trained neural network objects (defined by network parameters) are saved into a disk file and then read in by FirstWin (this process is called *serialization* in computer programming language). Based on the parameter values, FirstWin reconstructs the trained neural network objects for use in estimating the O-D travel time. In the routing and scheduling process, the neural network acts similar to a look-up table that gives the O-D travel time based on input data including location coordinates and departure time at the origin.

The performance of the ANN model as compared with the other estimation methods was tested by Fu and Rilett.<sup>5</sup> It was found that the ANN-based method significantly outperforms the distance-based method and is computationally efficient enough to be used in the routing and scheduling process for solving large-sized problems (over 1000 trips).



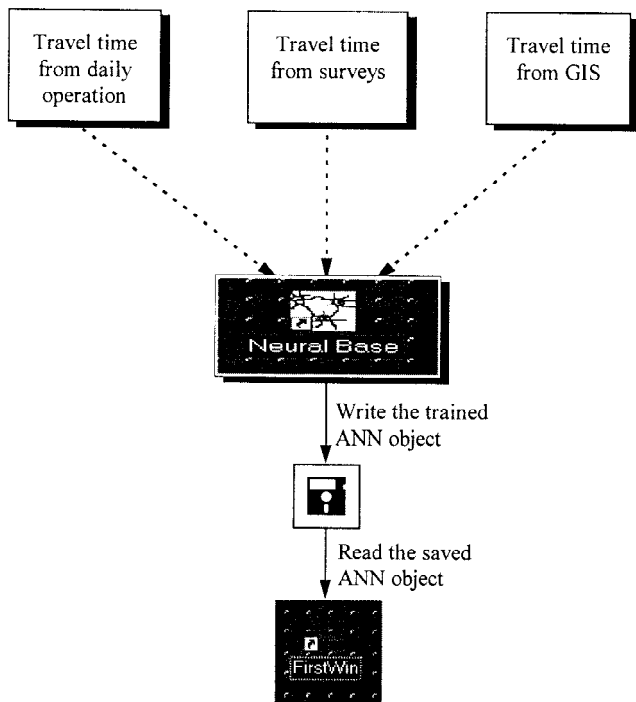


Fig. 6. Integration of FirstWin and NeuralBase.

## 4 SYSTEM IMPLEMENTATION

### 4.1 Hardware and software environment

FirstWin is written in Visual C++ using an object-oriented programming (OOP) technique that provides the flexibility to easily modify and extend the program as the system is continuously validated and enhanced. For example, the program was developed initially to handle only deterministic travel time. It was later extended to use a stochastic travel time model simply by adding a new class (representing stochastic travel time) derived from the old travel time class (deterministic travel time) without modifying the original scheduling algorithms. With OOP diagrams, the algorithms developed allow handling of future travel time models that have yet to be developed.

Because the input-output of the software is built on Microsoft's ODBC standard, a DBMS that supports ODBC is required. FirstWin is compiled as a 32-bit program run on Microsoft Windows 95 and Windows NT platforms. The size of the problem (e.g., number of trips and vehicles) and number of zones used for O-D travel time estimation are not restricted in the program but by the available computer memory.

### 4.2 Interface for data input, output, and editing

A graphic user interface (GUI) is used to display all system-status data, including trips, vehicles, travel time, scheduling

parameters, schedules, and various measures of effectiveness (MOE). FirstWin imports and exports data using the ODBC database engine, which allows the user to work on the same database application they are using and therefore requires much less data-preparation effort. Although most of the data are prepared outside the program and input from a database, an interface is set so that the user can easily modify the data, if necessary.

FirstWin also provides an interactive interface that allows schedulers to manually preassign trips to certain vehicles, to modify the schedules provided by the automated scheduling process, and to control which vehicles are to be used and which trips are to be scheduled.

### 4.3 Communication with other programs

FirstWin has been developed based on Microsoft's Object Linking and Embedding (OLE) technology and supports OLE automation. This means that it can be integrated easily with other application programs as a client-server system. The other applications can run on either the same or different computers, which may communicate with FirstWin through networks such as a local-area network (LAN) or the Internet. With this type of computing technology, FirstWin can be used as a real-time scheduling server to provide real-time scheduling functions such as reoptimizing schedules with new requests or trip cancellations. In this case, an instance of FirstWin may be kept running on a server computer, updating all the system data including working schedules, current vehicle locations, and trip information based on real-time collected data. The optimization routines in FirstWin can be called, for example, by a program at the dispatching computer to insert new trips to the existing schedules.

## 5 COMPUTATIONAL EXPERIENCE

This section illustrates the performance and computational efficiency of FirstWin in solving real application problems for both on-line and off-line operations. The data used for this demonstration were generated from the operational data for a typical day of the Disabled Adult Transportation System (DATS) in the city of Edmonton, Alberta, Canada. Edmonton, with a population of over 600,000, is also served by a conventional bus system and a light rail transit line. The regional population is almost 900,000.

The operational setting includes a total of 3024 trips and 109 vehicles with a designated service time from 6:00 A.M. to 10:00 P.M. The basic input data and scheduling parameters used are summarized in Table 1. All runs were performed on a Pentium 90 IBM-compatible PC with RAM of 24 MB.

The O-D travel time estimation is zone based with a total of 515 zones. Three travel time models are considered: constant, dynamic, and both dynamic and stochastic. Because only a

**Table 1**

Test problem description and scheduling parameters

<i>Test problem</i>		
Vehicles	Sedans (4 passenger seats; regular seat only)	33
	Minivans (6 passenger seats; regular seat only)	16
	Large vans (over 7 passenger seats; regular set and wheel chair)	60
Trips	Single trips (one to one)	2413
	Group trips (one to many, many to one and many to many)	632
<i>Scheduling parameters</i>		
	Maximum allowable ride time (minutes)	90
	Maximum allowable deviation from the most desired pickup and drop-off time (minutes)	30
	Maximum no. of trips per vehicle (trips)	40
	Initial no. of vehicles	100
	Minimum allowable probability	70%

**Table 2**

O-D travel time model and data

<i>Case</i>	<i>Description</i>
I	<ul style="list-style-type: none"> <li>• Constant travel time: Travel time for each O-D pair is a single value (mean)</li> <li>• A.M. peak travel time is used.</li> </ul>
II	<ul style="list-style-type: none"> <li>• Dynamic travel time: Travel times for each O-D pair are multiple values (mean) depending on time of day.</li> <li>• Three time periods: A.M. peak (7:00 A.M.–9:00 A.M.), P.M. peak (4:00 P.M.–6:00 P.M.) and off-peak (other).</li> <li>• For each O-D pair, travel time during A.M. peak period is assumed to be the same as the A.M. peak, and travel time during the off-peak period is 60% of the travel time during the A.M. peak.</li> </ul>
III	<ul style="list-style-type: none"> <li>• Dynamic and stochastic travel time: Travel times for each O-D pair are multiple values (mean and standard deviation) depending on time of day for each O-D pair.</li> <li>• Three time periods: Same as case II.</li> <li>• The mean travel time for each O-D pair is determined as in case II.</li> <li>• Variance of travel time for each O-D pair is calculated based on a constant variance to mean ratio of 0.2 minutes.</li> </ul>

database of average A.M. peak travel time was available, the travel times used for this analysis were created to represent these three travel time patterns, as described in Table 2.

The scheduling results obtained by FirstWin for three different cases are summarized in Table 3. The MOE values listed in Table 3 show that the use of different travel time models may result in significantly different schedules. As the results from case II and case III imply, a more realistic model (case III) requires more vehicles to service the same number of trips compared with using a less realistic model (case II). This is expected because in case III a reliability of 70% is guaranteed for all schedules generated, as compared with 50% in case II. The total vehicle travel time and vehicle productivity are quite similar for these two cases, which indicates that system productivity can be maintained, although more reliable service can be provided when using a more realistic model.

It should be noted that the extent to which a more realistic model provides an advantage over a less realistic model would need to be further determined by an empirical investigation. However, it can be anticipated that using more realistic travel times will yield schedules that are more efficient and more easily followed by drivers.

Table 3 also shows the relatively long computational time required to schedule all trips at once for each case (about 1 hour). However, this amount of time is acceptable when

**Table 3**

Summary of scheduling results

<i>MOE</i>	<i>Case I</i>	<i>Case II</i>	<i>Case III</i>
Vehicles scheduled	95	96	106
Trips scheduled	All	All	All
Total vehicle hours	772.0	551.72	581.43
Average ride time (min)	3.94	5.52	5.24
Average deviation from desired pickup/drop-off time (min)	38.73	31.63	27.89
CPU time for scheduling all trips at once (s)	2956	3365	3644
CPU time for insert of an extra trip after all trips are scheduled (s)	0.8	1.0	1.5

compared with the time available for off-line scheduling (e.g., 24 hours in advance).

After all trips are scheduled, the algorithm was run with the task of inserting a new trip to simulate an on-line operation with the assumption that the existing schedules are to be maintained. It can be seen that the CPU time required in all cases is less than 2 seconds. These computational performances imply that the program is satisfactory for both off-

line and real-time scheduling situations when a more realistic travel time model is used.

## 6 CONCLUSIONS

This article presented an overview of the concepts, models, and computational technologies used in a new vehicle routing and scheduling software called FirstWin, which has been developed for dial-a-ride paratransit operations. The program was conceived ultimately to be used in the Automated Dial-a-ride Paratransit Operation Systems (ADPOS) conceptualized in this article. Compared with existing tools, the software system has the following advantages<sup>10,14</sup>:

- It includes a new DARP model that explicitly considers the dynamic and stochastic nature of O-D travel times. This model allows adequate use of the large amount of data available due to ITS technologies to improve the efficiency and reliability of vehicle service.
- A new O-D travel time estimation method based on ANN technique is incorporated that can significantly improve travel time estimation while being computationally efficient enough to be used in the routing and scheduling process for solving large-sized problems.
- A user-friendly graphic interface and a database-oriented data input-output scheme are used so that data preparation and use of the program are much easier. OOP diagram and OLE automation support make the program more extendible and more easily integrated with other application programs for real-time operations.

A computational analysis using real application problems has shown that the program is efficient enough to be used to provide both on-line and off-line scheduling services. A practical version has been tested extensively and will soon be used regularly to schedule more than 3000 daily trips by the Disabled Adult Transportation System (DATS) of the city of Edmonton.<sup>4</sup>

Finally, it should be noted that although the program was developed specifically for dial-a-ride paratransit system, the methodology and algorithms can be modified and extended for use in other applications such as school bus systems and food and package delivery systems.

## ACKNOWLEDGMENTS

This research was supported in part by the Natural Sciences and Engineering Research Council of Canada. We are grateful to the Disabled Adult Transportation System of the city of Edmonton for its cooperation in providing the test data and for useful suggestions to the initial FirstWin versions. We

would like to thank the editor and three anonymous referees for their valuable comments and suggestions.

## REFERENCES

1. Bodin, L., Golden, B., Assad, A. & Ball, M., Routing and scheduling of vehicles and crews: The state of art, *Computers and Operations Research*, **10** (1983), 69–211.
2. Dial, B. R., Autonomous dial-a-ride transit: Introductory overview, *Transportation Research C*, **3**(5) (1995), 261–75.
3. Fu, L., Real-time vehicle routing and scheduling in dynamic and stochastic traffic networks, Ph.D. dissertation, University of Alberta, Edmonton, Alberta, Canada, 1996.
4. Fu, L., User's guide: FirstWin—A tool for routing and scheduling dial-a-ride paratransit vehicles (draft), TransOptimum, Edmonton, Alberta, Canada, 1996.
5. Fu, L. & Rilett, L. R., Dynamic O-D travel time estimation by using artificial neural network, in *Proceedings of Vehicle Information and Navigation System (VINS) 1995 Annual Meeting*, Seattle, Washington, 1995.
6. Fu, L. & Teply, S., Improving the reliability of paratransit service using better travel time information, in *Proceedings of the Joint Conference of the Canadian Institute of Transportation Engineers*, Pacific Northwest ITE Quad Section and Western Canada Traffic Association, Vancouver, British Columbia, Canada, April 13–16, 1997.
7. Golden, B. L. & Ball, M., Shortest paths with euclidean distance: An explanatory model, *Networks*, **8** (1978), 297–314.
8. Jaw, J., Odoni, A. R., Psaraftis, H. N. & Wilson, N. H. M., A heuristic algorithm for the multivehicle advance request dial-a-ride problem with time windows, *Transportation research B*, **20B**(3) (1986), 243–57.
9. Keeney, R. L. & Raiffa, H., *Decisions with Multiple Objectives*, Wiley, New York, 1976.
10. Lave, E. R. & Piras, P., *A Handbook for Acquiring Demand-Responsive Transit Software*, TCRP Report 18, Transportation Research Board, Washington, DC, 1996.
11. Mohammadi, R., Journey time variability in the London area: 2. Factors affecting journey time variability, *Traffic Engineering and Control*, **38**(6) (1997), 337–46.
12. Mullen, D. & Fu, L., Travel time study in the city of Edmonton, unpublished report, Department of Civil and Environmental Engineering, University of Alberta, Edmonton, Canada, 1996.
13. Savelsberch, W. P. M. & Sol, M., The general pickup and delivery problem, *Transportation Science*, **29**(1) (1995), 17–29.
14. Stone, J. R., Nalevanko, A. & Tsai, J., Assessment of software for computerized paratransit operations, *Transportation Research Record* 1378, Transportation Research Board, Washington, DC, 1993, pp. 1–9.
15. Wilson, N. H. M. & Colvin, N. H., Computer control of the Rochester dial-a-Ride system, report R77-31, Department of Civil Engineering, M.I.T., Cambridge, MA, 1977.
16. Wilson, N. H. M. & Weissberg, H., Advanced dial-a-ride algorithms research project: Final report, report R76-20, Department of Civil Engineering, M.I.T., Cambridge, MA, 1976.