SIMULATION OF A COMPUTER AIDED ROUTING SYSTEM (CARS)

N. H. M. Wilson  
Massachusetts Institute of Technology

J. M. Sussman  
Massachusetts Institute of Technology

L. A. Goodman  
MITRE Corp.

B. T. Higonnet  
Massachusetts Institute of Technology

Abstract

CARS is a system designed to provide a taxi-like service at a mass transit-like cost. It allows potential passengers to request service from their homes via telephone, with calls being processed by a central computer facility. The computer periodically executes a routing algorithm which assigns vehicles to passengers and communicates this routing information to the vehicles. The system is 'real time' in that it will pick up a passenger and shortly after a request will deliver him to his destination within a guaranteed time (with a minimum number of deviations for collecting and delivering other passengers). The key to CARS is the routing algorithm. Since labor and vehicular costs are a major portion of the total system cost, an algorithm is required which can provide an effective dynamic service with a minimum number of vehicles. A variety of such algorithms have been proposed, but these algorithms can not be evaluated in an analytic fashion. Hence, a comprehensive simulation model has been developed to test and compare these routing algorithms. The model facilitates the investigation of relationships between such parameters as number of vehicles and quality of service. This paper describes the methodology of the simulation model and the economic gains (in terms of the need for fewer vehicles) realized through its use. The use of an ARDS storage tube display to produce graphical output from the model is also discussed.

INTRODUCTION

CARS (Computer Aided Routing System) is a new concept in public transportation which is intended to provide door-to-door transportation service at a cost close to that of existing mass transit. The research and design of the system is being carried out by several academic departments and the special laboratories of M.I.T. This work was initially sponsored through the U.S. Department of Housing and Urban Development and is currently being supported by the U.S. Department of Transportation.

* In more advanced versions of the system, a user will be able to touch tone his message, thus negating the need for human operators.

About 80 people at M.I.T. are involved in the project. Initial research demonstrated the economic and technical feasibility of the CARS concept and the current effort is aimed at the design and implementation of a prototype CARS system.

THE SYSTEM

CARS works basically as follows: customers call for service from their homes (or some other origins) using a telephone. These calls for service are received by operators* who input the required
information (e.g. origin and destination of trip, number of passengers, perhaps a billing number) into a digital computer. The computer is concerned with assigning a CARS vehicle to pick up and deliver each new passenger. This computer also automatically generates appropriate messages containing new assignments for the various vehicles. These messages are transmitted in digital form to the vehicle. Each vehicle has an encoder-decoder on board which

1. translates the received digital messages into printed form for the driver and

2. allows the driver to send messages back to the central computer.

When a vehicle arrives to pick up a customer, the driver pushes a button which prompts the computer to send him information on his next stop. An emergency voice channel will also be provided for the driver. The system and information flows are shown schematically in Figure 1.

A customer will be picked up within 10 to 15 minutes of his initial call for service and once picked up will be taken reasonably directly to his destination. The customer will not generally be taken straight to his destination because the vehicle will have to make diversions to pick up and deliver other customers. Hence the service is very much like a shared taxi service: it is a centrally controlled door-to-door system which dynamically responds to requests and allows efficient sharing of vehicles. The design goal is that the ratio of the time spent in the system (i.e. elapsed time from phone call to delivery at destination) to direct driving time in a private automobile should average at most 2.5, and should in no case exceed 3.0.

In order to provide the taxi-like service at mass transit fare levels (say, 50¢/trip), the operating and capital costs must be kept as low as possible. Cost analyses indicate the most important cost components are the vehicles and the associated driver salaries. In some cases, these components constitute up to 75% of total system cost. Thus, one obvious way to keep the system cost low is to reduce the number of vehicles. It is here that the computer plays its most important role.

The hypothesis is that the computer can perform the assignment of new demands to vehicles very efficiently, in particular much more so than could a human multi-dispatcher system. Thus, savings in terms of number of vehicles (and associated drivers) are realized when a computer is introduced. Analysis shows the cost of the computer is far less than the savings in vehicle related costs that are realizable by using a computer to perform the scheduling function. Put another way, a very good but extremely expensive transportation service could be offered by using many vehicles with, in the limit, a one passenger per vehicle system (or a standard taxi service). However, almost the same levels can be provided at lower cost by using fewer vehicles.
and coordinating them by computer.

Of course, having a computer to do scheduling is no solution unless an efficient algorithm to perform the customer-to-vehicle assignment exists. The development of such algorithms is one of the main tasks of the CARS research effort. These algorithms must work in an extremely complex and difficult environment which is stochastic in several respects. The demands, which appear randomly in time, have origins distributed probabilistically in geographic space. To compound the problem, the algorithm must be performed in 'real time'. All these factors plus further research have lead to the conclusion that classical optimization solutions to the customer-to-vehicle assignment problem are not feasible. (1) While theoretically they could be formulated and solved for small problems, no reliable method for solving problems of realistic size has been found. This statement is true even if the real time constraints under which the algorithm must operate are considerably relaxed.

Therefore the algorithms that have been developed are heuristic in nature. When a new demand arises, the algorithm uses some heuristic rule to assign the demand to a particular vehicle and to appropriately insert the new origin and destination in the vehicle's route. At all times each vehicle has a provisional route (i.e. sequence of stops) associated with it. Provisional routes are updated with the occurrence of new demands and a vehicle is not committed to a particular stop until the last possible moment.

Heuristics may be simple or complex and this paper will not dwell on the heuristic rules themselves. (1) There is a clear need for a means of evaluating and comparing different heuristics. Certainly no analytic model could predict the effectiveness of the heuristics, for much the same reasons as given for the failure of classical optimization techniques. Therefore, to help in the evaluation of the various heuristics, a simulation model for the CARS algorithms was developed. The remainder of this paper describes this simulation model and the results obtained from it.

THE MODEL

The model was implemented in FORTRAN and is an event structured formulation. It is composed of 40 subprograms, and is approximately 100,000 bytes in length (including data areas). Approximately two man-years of effort were required to design, develop, and test the model. Since its completion, about the same effort has been expended in exercising the model and in analyzing its results. Before work started on the development of the model described in this paper, several simpler CARS models were developed (10,4). Thus, the staff was reasonably experienced with the issues before this effort began in earnest. While the development of this progress of models was happenstance (the original funding was small and hence the effort was low-key) it is felt in retrospect that it was an efficient way to develop the model described here. In effect, an incremental approach was adopted. Rather than begin with a fullblown model, simpler models were developed first. The final model is considerably better as a result of this and the approach is certainly recommended.

The question 'why FORTRAN?' might well be asked. At the outset, it should be noted that FORTRAN and GPSS were the only languages easily available to M.I.T. at the time of the model's development. FORTRAN was chosen by default since it was felt that GPSS did not lend itself to this model. The algorithms being modelled typically required a great deal of algebraic computation and since the use of GPSS in such an environment is less than optimal, FORTRAN was chosen. Some evidence that this was the correct step exists. One of the authors (in a sadistic moment) assigned an extremely simplified subset of the CARS model (e.g. all customers going to the same destination) to
A graduate class in simulation techniques with the proviso that GPSS be used. Even for the 'easy' formulation, GPSS proved to be a difficult language with which to work. In fact, very few students produced a working model.

As mentioned above, the large number of complex algebraic operations led to a choice of a compiler based language (FORTRAN) rather than an interpretive system (GPSS). While the compiler approach is certainly sound, it is felt that FORTRAN is not the best language for this application. Specifically, the CARS model makes heavy use of complex data structures, especially list structures, e.g. lists of demands, lists of vehicles, etc. A language with good list processing facilities (which FORTRAN is decidedly not) would have simplified the modelling task immensely. For example, SIMSCRIPT with its good data structuring capabilities might have made for easier implementation. However, SIMSCRIPT was not available and hence the choice was, in effect, made independently of these considerations.

The simulation model was designed primarily to allow the analyst to test and evaluate various heuristic rules. These tests were to take place under a wide variety of exogenous conditions. With this in mind, the model was constructed so that the analyst could easily implement the heuristic of current interest either by varying an input parameter or by making simple changes in the code of the model itself. This latter capability is realized by writing the model in a rather fine grained modular fashion, thus allowing changes to be made in a straight-forward way. (It should be pointed out that while modularity allows changes to be implemented easily, one does pay a price in terms of efficiency of execution. This is basically the higher cost of general, flexible structures when compared to special purpose, inflexible formulations).

In addition, the input to the model allows for the definition of a great many CARS environments. The following is a partial list of inputs and is given merely as an indication of the flexibility of the model.

1. Number of vehicles
2. Vehicle capacity
3. Average vehicle velocity
4. Time distribution of demands (chosen from a predefined family or defined by the analyst)
5. Spatial distribution of origins and destinations (uniform or non-uniform)
6. Intermediate output intervals
7. Output Options

Considerable effort was expended in the model's design so as to provide the user a great deal of flexibility in his selection of random number strings. First, he can specify a separate seed for each of the random processes in the model. Second, he can specify whether the random number string or its complement string (obtained by subtracting the original number from unity) should be used in the simulation. These kinds of control over the pseudo-random processes in a simulation allow the user to do two things.

First, he can parametrically vary the controllable variables in the model (such as the heuristic rules themselves) while, in effect, eliminating unwanted randomness in the results. By using the same random number strings, he can duplicate the external features of the model from run to run while varying other parameters. In this way, positively correlated results can be obtained. This allows efficient (in terms of simulation run lengths) comparison of various strategies and also permits more rapid check-out and debugging when changes are made to the model.
Second, the analyst can achieve a complementary pair of runs by holding the parameters of the model constant and running the model twice, first with standard (uniform 0,1) strings of random numbers and second with one or more of the random strings complemented. In this way, it is possible to obtain pairs of negatively correlated runs. The variance of the results obtained by summing the results of the two runs and dividing by two is generally less than the variance of a single run of twice the length, and hence efficiency gains can be realized. The initial work in this area has been quite encouraging.

To summarize, the model allows the analyst to conveniently experiment with techniques of this sort by permitting flexible specification of the random number strings and hence of the probabilistic structure of the model.

Output is equally extensive with its volume easily controlled by the analyst. Its precise make-up is documented in reference[2].

MODEL RESULTS
As was explained in the preceding section, the model was designed to test and evaluate a variety of routing heuristics for possible use in a real CARS application. Its use was an essential component in estimating the economic feasibility of the system and, as such, was crucial to the research and development program. Before describing the experiments that were performed and the results obtained, it is important to understand the criteria for judging the effectiveness of a given algorithm.

The variables over which the algorithm exercises control are the cost to the user of the system and the service times (waiting time plus travel time) experienced by the users. The aim of an algorithm is to minimize both of these conflicting variables (lower service times are generally achieved by using more vehicles hence increasing the cost). Different algorithms should, therefore, be compared in this two-dimensional space of cost and time. An added complication is that the service times are random variables and different users of the system may be interested in different types of distributions. For instance, someone travelling to catch a plane is concerned primarily with the latest time he might arrive at the airport using this system, while a shopper might be much more interested in the mean service time. Clearly then, both the mean and extreme worst service times are relevant. Considering these factors, the term level of service has been defined as the mean value of the ratio of service time to direct driving time taken together with the extreme worst service time resulting from the operation of a particular heuristic. The better the service, the lower the level of service figure, though it can never be less than 1.0. Then the measure of effectiveness for a heuristic in a specific situation is the number of vehicles required to provide a given level of service.*

Experiments using the model have been run for two main purposes.

(1) To compare different heuristic rules in terms of the measure of effectiveness defined above.

(2) For a given heuristic, to predict how CARS would function in a variety of operating environments defined by varying the following parameters:
(a) Area size and shape
(b) Demand level
(c) Desired Service characteristics.

The general form of heuristic investigated involved attempting to insert the new demand's origin and destination in all possible positions on

*Using number of vehicles rather than system cost is a valid approximation since as described earlier the dominant cost components in CARS are vehicle related (including labor).
all vehicles' future routes. Associated with each attempt, a function \( Z \) is computed. This function measures the disruption caused by the insertion of the new origin and destination within a route. The insertion for which this function \( Z \) is minimized determines which vehicle will be assigned to the new demand. The route for that vehicle is then updated to reflect this.

In designing various heuristics two factors have to be considered:

1. Good service for current system users must be provided.
2. The heuristic should ensure that service for future users will not be jeopardized.

Considering just the current users, a possible selection criterion is:

\[
\text{Minimize } Z_1 = N_1 e_1 + N_2 e_2 + ST
\]

where \( N_1 \) is the number of customer deliveries after the insertion of the origin stop for the new demand (see Figure 2).

\( N_2 \) is the number of customer deliveries after the insertion of the destination stop for the new demand.

\( e_1 \) is the vehicle detour time due to the insertion of the origin of the demand.

\( e_2 \) is the vehicle detour time due to the insertion of the destination of the new demand.

\( ST \) is the service time of the new customer for this insertion.

Then \( Z_1 \) measures the time which will be lost by all passengers currently on the system if the new demand is assigned in this manner.

Similarly considering just the future demands, a possible selection criterion is:

\[
\text{Minimize } Z_2 = e_1 + e_2
\]

This frees all vehicles as soon as possible thus enabling them to service new requests.

Figure 2. One Possible Insertion of New Demand C

Figure 3 shows a comparison of these two criteria for a 3x3 mile area and for the same level of service. It is seen that the criterion \( Z_1 \) favoring current users gives better results though there is not very much difference.

Combining \( Z_1 \) and \( Z_2 \) produced the following criterion:

\[
\text{Minimize } Z_3 = (N_1 +1) e_1 + (N_2 +1) e_2 + ST
\]

Results of using \( Z_3 \) are also shown in Figure 3 and it is seen that very little improvement is obtained over \( Z_1 \). Of all heuristics tested to date, however, \( Z_3 \) produces the best results.

Figure 3 also illustrates one aspect of the second objective of the model experiments; the effect of parametric variation of the demand level on the cost of providing service. It is seen that for any of the investigated heuristic rules and for the same area and level of service, there is an approximately linear relationship between the number of vehicles required and the demand level.

However, the vehicle productivity (in terms of passenger trips per vehicle hour) does increase with increasing demand level as would be expected. This means that the more people who can be persuaded to use the service, the lower will be the unit cost.
Figure 4 demonstrates for the third (and best) heuristic rule the effect of operating in different area sizes at the

![Figure 3 Alternative Heuristics](image)

Figure 3 Alternative Heuristics

Figure 4 demonstrates for the third (and best) heuristic rule the effect of operating in different area sizes at the same demand rate per square mile. The curves refer to demand rates of 25 and 40 per square mile per hour in a 3x3 mile area. Here too it is seen that the relationship is approximately linear with more vehicles required to service larger areas. In this case however, the vehicle productivity remains approximately constant and independent over area size. This indicates that CARS does not stand to gain great economies of scale with respect to area size.

![Figure 4 Effect of Area Size](image)

Figure 4 Effect of Area Size

Figure 5 shows, again for the third heuristic, the variation of level of service with number of vehicles operating. The curves refer to 3x3 and 5x5 mile areas with demand rates of 25 per square mile per hour. The rectangular hyperbolic shape simply reflects the fact that a large number of vehicles can provide a good service whereas a small number can not. It is seen that in the 3x3 mile area, as the mean level of service approaches unity, the vehicle requirements increase rapidly while at lower levels of service, small changes in the number of operating vehicles can have a sizeable impact on the quality of service. These results indicate that CARS should aim at a level of service of about 2.5.(3)
Future developments include the simulation of particular geographical areas in which real street networks and expected demand patterns would be incorporated. A further major step forward would be the development of a complete CARS system simulation rather than the algorithm simulation which exists now. In the development of CARS, simulation is certain to play a vital role right up to (and indeed after) the first passengers are carried; in a complex stochastic system such as this, there is no alternative.

INTERACTIVE CHARACTERISTICS OF THE MODEL

Many authors have expounded on the advantages of working with computers in an interactive (e.g. time-sharing environment). (5,9) It is sufficient here to say that it is worth a great deal to have the ability to interact with the actual operation of the model, especially in a simulation and more especially in the case of CARS where the question 'what if this happened?' is always present. With these advantages in mind, the model was implemented and is currently operational in an interactive time-sharing system driven by an IBM 360/67 at M.I.T. This environment provides excellent development and testing aids and has unquestionably accelerated the research.

One very important way in which operating interactively was helpful was that the use of computer graphics became feasible. CARS is a classic example of the application of graphics. With demands arising randomly in space and vehicles moving about a large area, an on-line display is an extremely important aid to the intuition of the analyst. The mere printing out of results such as mean waiting time, for example, could easily mask the fact that the heuristic is generating pathological tours (e.g. around the block three times) which would be unacceptable to customers. Graphic display, on the other hand, easily uncovers these abnormalities. In general, the analyst using

CONCLUSIONS AND FUTURE DEVELOPMENTS

The model proved invaluable in demonstrating certain characteristics of the CARS system and justifying continued work to elaborate the concept.

In particular it helped validate algorithmic ideas and indicate how the system might behave in a wide range of environments with variation of area size, demand level, and service characteristics. The model was used successfully to compare algorithmic choices and in this way helped to clarify to the analyst the operation of the system. The best heuristic based selection on both future and current system users with more weight on the latter. It was found that the system was not very sensitive to differences between several rational heuristics. The service should be more efficient at higher demand rates but large service areas should probably be avoided.
interactive graphics can ensure that the model's behavior 'makes sense' without preprogramming extensive logic checks. In short, the task of model validation is made much simpler if graphics output is available. The model is designed to display, at pre-determined but easily modified intervals, the most relevant aspect of the simulation as it has developed up to that point (i.e. display a passenger, display all passengers who have been waiting for t or more minutes, display vehicles positions, display vehicle x's route, display the route of the vehicle to which passenger y is assigned, display system performance statistics such as average waiting and travel times, etc.).

There are two important points to be made about these display functions. First, it is very difficult to predict the best format for the display features. In fact, almost all display features had to be changed after we gained some experience in their use because they were clumsy or not sufficiently informative.

Second, hardware availability constrained us to use an ARDS (Advanced Remote Display Station) storage tube display (as opposed to a refreshed display such as the IBM 2250) and in spite of our efforts to predict what would be a meaningful display, the tube had a remarkable propensity to becoming cluttered. Hence the display is designed to be under the complete control of the user. In what are felt to be meaningful increments, the viewer displays and erases at his discretion: relatively little is done automatically by the display routine itself. The screen does however become cluttered if a large simulation is run.

In short, it became possible for the analyst to monitor the progress of the simulation and algorithm performance during the running of the simulation, and this in considerable detail. Possession of such knowledge as soon as the events involved took place opened up the prospect of interactive communication between the analyst and the simulation model. Because the viewer had a much more detailed and much more intuitive picture of the model during the run, he was now in a position to meaningfully affect its future course.

For instance, the analyst, given a particular situation in terms of vehicle and passenger positions, might want to see what would happen if a passenger originating 'here' wanted to go 'there'. Because of the graphical output and the ability to input through the display tube, the analyst can easily specify origins and destinations for the hypothesized new demand.

In addition to being able to manually specify a demand, the analyst can ask to see the choice of the current algorithm for that demand, before the model is committed to that choice. Thus if the algorithm wanted to assign vehicle x to passenger y, the analyst would be informed and would have the opportunity to over-ride the algorithm's choice. That is, he could prevent the desired assignment and force the algorithm to use the second best choice. In particular, by simply repeating this procedure, the analyst can see, in decreasing order of desirability, all the algorithm's choices and then indicate to the model which should in fact be used.

With the above capabilities in hand, the analyst truly has a new level of information available to him in readily understood form which permits him to delve in various aspects of algorithm performance. Since the display and interactive aspects of the model are under his control, he does not risk being deluged with more information than he can handle.

Future developments for the graphics additions to the model consist primarily of automatic entry into the graphic interactive mode whenever a specified set of conditions occurs (e.g. whenever mean level of service exceeds a given value, etc.). It is not known how the human engineering of such a feature will work out, but it should be
possible to avoid too much automatic graphical output and thus keep the display useful.

In order to indicate the range and utility of the graphic display, photographs taken during an actual run (3x3 mile area, 100 demands/hour, 12 vehicles) are described below. Figures (6) through (8) refer to time 45 minutes into the simulation run. Figures (9) through (15) refer to time 60 minutes and shortly thereafter.

Figure (6) shows all people who have requested service but have not yet been collected. Each dashed line joins the origin and destination for one of these waiting users. Associated with each user is a passenger identifier (the 4-digit numbers in the figure) and the time at which the service request was made (the 3-digit numbers).

![Figure 6](image)

Figure (7) shows the current projected route of vehicle 11. Next to each stop on the route is the corresponding passenger identifier, the letter 'P' or 'D' indicating pickup or delivery respectively, and the predicted arrival time at that stop. The vehicle is shown at its last reported stop, whereas it is in fact travelling to its first stop. Its projected route is to deliver demand 43 (currently on board), pickup demand 81, pickup demand 77, deliver demand 77, and finally deliver demand 81. Then, according to current plans, the vehicle will be empty and unassigned at time 65 minutes.

![Figure 7](image)

Figure (8) shows the activity of vehicle 11 in the immediate past. The labelling is the same as in the previous figure with the times being the actual arrival times.

![Figure 8](image)
Figure (9) at time 60 minutes shows all passengers waiting to be picked up.

Continuing to follow vehicle 11, Figures (10) and (11) show its current projected route and recent history, respectively. Comparing these figures with Figure (6) it is seen that more stops have been added to vehicle 11's route since time 45. Specifically, passenger 96 has already been collected and the vehicle is now also scheduled to serve passenger 104. Consequently the expected delivery times of passengers 77 and 81 have been revised. This is, of course, a typical example of the system's dynamic response to changes in the demand state.

Figures (12) through (15) relate to the interactive aspect of the graphics in the model. In Figure (12) passenger 111 has been manually input through the screen (much as one would use a light-pen) with the desired origin and destination points. Also shown is the assignment which is chosen as best by the heuristic. If the analyst did not interfere, the new passenger would be assigned to vehicle 8 and (as indicated by the underlining) would be collected just before passenger 103 was to be delivered.
Figures (13) and (14) show the heuristic's second and third choice of assignments, respectively. Finally, Figure (15) shows the revised route of vehicle 11 (the heuristic's second choice) after passenger 111 has been assigned to it at the analyst's behest (at time 64).
This last sequence of figures demonstrates how the analyst can obtain an excellent idea of how the heuristic is actually functioning. In this way errors in both theory and implementation can readily be detected and the insight of the analyst is deepened so that better heuristics can be developed.

REFERENCES


BIographies

Nigel Wilson, Assistant Professor in Civil Engineering at M.I.T., received his SB from the University of London in June 1965, his SM in January 1967 from M.I.T., and his Doctorate in Civil Engineering in October 1969 from M.I.T.

Joseph Sussman, Assistant Professor in Civil Engineering at M.I.T., received his B.S.E. from City College of New York in January 1961, his M.C.E. from the University of New Hampshire in June 1963, and his Doctorate in Civil Engineering from M.I.T. in January 1968, and is also Senior Consultant and Director, Engineering Computer International Inc.

Laurence A. Goodman, currently with MITRE Corporation, received his SB in June 1967, an SM in November 1968 both from M.I.T.

Trevor Higonnet, Research Engineer at M.I.T., received his SB in June 1965 and his MS in June 1966 both from Harvard.