Proceedings of the 1987 Winter Simulation Conference
A. Thesen, H. Grant, W. David Kelton (eds.)

A SIMULATION ANALYSIS OF DEMAND AND FLEET SIZE EFFECTS ON TAXICAB SERVICE RATES

William A. Bailey, Jr.
University of North Carolina at Charlotte
Charlotte, NC 28223
Thomas D. Clark, Jr.
The Florida State University
Tallahassee, FL 32306-1042

ABSTRACT

The research presented in this paper was directed at understanding the dynamic interaction between demand, service rate, and policy alternatives in a typical urban taxi system. The general nature of such systems is discussed and a SLAM II simulation model of the system introduced. A series of experiments with the model indicate that customer waiting time is relatively insensitive to changes in demand but highly sensitive to changes in the number of taxi cabs.

1. INTRODUCTION

Taxicabs provide an important transportation service to the people of many communities across the country, especially those living in urban areas. A number of complex, interacting forces exist in the management of a taxi system making management of its operation difficult. Very little research exists that has attempted to address the special structure and behavior of these systems.

Taxicabs constitute a form of transportation called paratransit. Kirby and his colleagues (1974) define it as "those forms of intraurban passenger transportation which are available to the public, are distinct from conventional transit (scheduled bus and rail), and can operate over the highway and street system." It includes those types of transportation between the private automobile and conventional mass transit, such as car pools, rental cars, dial-a-ride vehicles, jitneys, limousines and of course, taxicabs.

Paratransit can be conveniently grouped into three categories: those you hire and drive, those you hail or phone, and those for which you make prior arrangements with other travelers. Hail or phone service is often called demand-responsive transportation (DRT) and includes taxicabs, dial-a-ride vehicles and jitneys.

The operation of a dial-a-ride system has often been classified as a vehicle routing and scheduling problem. A routing and scheduling system deals with a collection of entities requiring service. If the entities to be serviced have no temporal restrictions and there are no precedence relations among these entities, then it is a routing problem. If each entity has a definitive service time, then a scheduling problem results. Otherwise, one is dealing with a combined routing and scheduling problem. Generally, routing and scheduling problems involve both task precedence relations and time window constraints. Task precedence relationships force the pickup activity for a task to precede the delivery activity for the task, and the pickup and delivery tasks must be with the same vehicle. The dial-a-ride dispatching problem clearly falls in the latter category.

Bodin and others (1983) have provided the most complete and comprehensive review of the routing and scheduling literature. The area of dial-a-ride routing and scheduling has received considerable attention. For research concerning problems with the static dial-a-ride problem, see Sexton (1979), Sexton and Bodin (1980), Baker (1981), and Psaraftis (1983). These reports deal with the single vehicle case, and Bodin and Sexton (1982), Jaw (1984), and Jaw and others (1986) deal with the multiple vehicle case. The objective in both cases is the minimization of customer inconvenience, which is usually some combination of excess ride time and desired pickup or delivery time deviation.

In the demand-responsive system, new customers must be inserted into the current routes and schedules. The problem is to determine the assignment of the new customer to a vehicle and then to recompute the new route and schedule for the vehicle to which the customer is assigned. Because this assignment is carried out in real time, the algorithms are heuristic and very rapid. Although Psaraftis (1980) has developed an exact dynamic programming solution to the single vehicle case, most realistic applications would be for much more complex multiple vehicle situations. Virtually all of the research in this area has been done by Wilson and his colleagues at MIT. This research includes Wilson (1967), Wilson and others (1970, 1971, 1976), Wilson and Higonnet (1973), Wilson and Miller (1977, 1979), and Wilson and Colvin (1977). Approximate analytic queuing models also have been developed by Hendrickson (1978) and Daganzo (1978) and a shared-ride taxi model has been advanced by Simpson (1978).

The operations of an exclusive-ride taxi system is not as easily classified as a scheduling and routing problem. The basic output of all routing and scheduling systems is essentially the same: for each vehicle a route and schedule is provided. Generally, the route specifies the sequence of locations to be visited and the schedule identifies the times at which the activities at these locations are to be carried out (Bodin and others, 1983). With SRT systems (dial-a-ride), a collection of customer origins and destinations always exists, forming (though only temporarily in the dynamic case) a transportation network. In demand-responsive exclusive-ride systems (taxis), customer origins are not known in advance and are served one at a time, thus no netowrk exists with which to apply mathematical programming techniques.

Taxicab dispatching seems to better be represented as an M/G/s queuing problem which is one where calls arrive according to a Poisson distribution (M), calls are serviced in accordance with some general service distribution (G), and there are s parallel servers (taxis) (Gross and Harris, 1985). The complexity of the service distribution, however, makes the problem too difficult to address using existing queuing theory unless very restrictive assumptions are made.

Even if the service distribution coule be derived, exact queuing formulae have not yet been developed for the M/G/s problem, although several approximations exist (Boxma and others, 1979; Hokstad, 1978; Kimura, 1983; Takahaski, 1977; Tijms and others, 1981; Yao, 1985). Nonetheless, queuing concepts do provide a useful framework for discussing the problem although meaningful analysis appears dependent upon the use of a computer simulation approach.

A coherent, continuing stream of research that specifically addresses the operation of a taxi fleet has not materialized as it has for other consumer transportation systems. This probably is a result of the taxi industry being fragmented and essentially private causing it to be overlooked by the government as a viable means of public transit. Little public funding for research has been provided for taxis, but in contrast, research on more publicly controlled modes of transportation, such as dial-a-ride, school busses, and mass transit busses has received considerable attention and funding, often from the U.S. Department of Transportation (DOT), Urban Mass Transportation Administration, under the Urban Mass Transportation Act of 1964.

Only three pieces of research specifically addressing the operation of an exclusive-ride taxi fleet were identified. These are Meyer and Wolfe (1961), McLeod (1972) and Gerrard (1974). All of these attempt to solve the problem via the analytical/queuing approach by making certain restrictive assumptions, although, in two cases, simulation was used as a validation tool. All three studies recommended the use of computer simulation rather than analytical methods to approach the analysis of taxi systems.

The difficulty in reaching analytical solutions to such problems lies in their inherent complexity. The infrastructure of the area, weather, the number of vehicles, demand, and the skill of drivers and dispatchers interact to produce a service rate that yields some level of profit for the taxi company. The dispatcher's job is primarily an "art" strongly dependent on experience and familiarity with the demographics and infrastructure of the area. In a system with a limited number of taxis in a small to midsize community, this may not be an unmanageable operational level problem for a gifted and experienced dispatcher, but there clearly is a point at which no dispatcher can maintain a grasp of the problem and still manage an efficient operation.

2. SYSTEM AND MODEL STRUCTURE

A conceptual model of the structure of the taxi system modeling environment is presented in Figure 1. This particular representation focuses on the variable structure of the system and not on the inherent complexity involved in a detailed taxi system model that captures the interactions among the variables and the resulting state conditions.

There are four sets of variables. The output or goal variables describe the two objectives of the taxi system: to maximize profitability and service quality. These two goals interact as improvements in quality raise costs and thus reduce profitability. The policy variables specify the means available to control the operation of the taxi fleet. These variables can be manipulated in order to identify the "optimal" combination with respect to the performance measures (goal variables). The environmental variables identify the various conditions under which the policies will be tested. The "optimal" policy mix must be determined under a variety of demand and cost scenarios. Finally, the parameters describe the constants under which the taxi model will operate. These values do not change but provide the necessary infrastructure in order to experiment with the model.

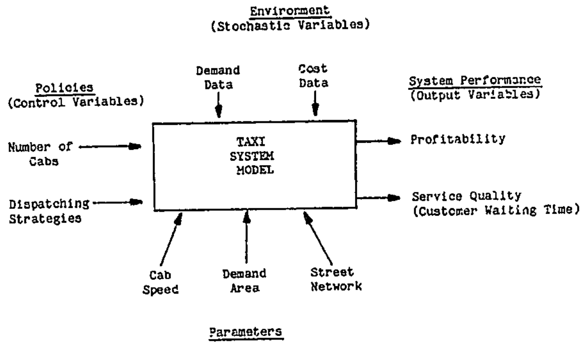

Figure 1: Variable Structure

The procedural structure of a SLAM II (Pritsker, 1986) simulation model of a representative taxi system is presented in Figure 2. The model uses a discrete event world view implemented in a SLAM network supplemented with extensive user written FORTRAN subroutines. Essentially, the cabs are moved around a city divided into twenty-five demand location zones. The extensive bookkeeping required also is implemented with user written routines. A discussion of the specific operation of the model and a documented version of it are available upon request.

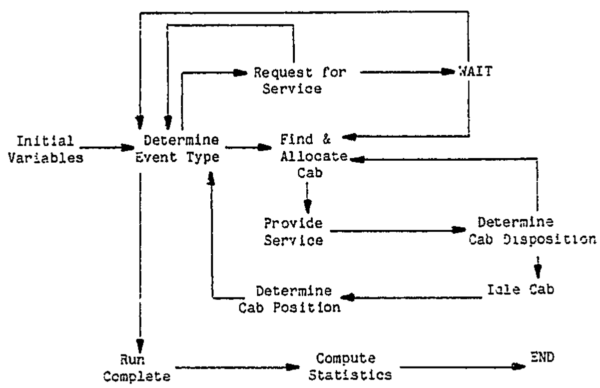

Figure 2: Overview of Model Structure

The variable of interest addressed in this paper is the second goal variable, service quality, which is represented in the model by a measurement of customer waiting time. As with most queuing systems, waiting time is an important measure of performance for the taxi system. The state diagram in Figure 3 more clearly illustrates the customer waiting time structure.

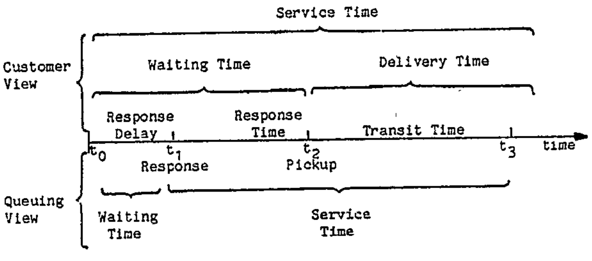

Figure 3: System State Diagram

Three states exist in the system as taxi service is provided. A customer calls at time t_0. If cabs are busy, then there is a delay until time t_1 (a response delay of $t_1 - t_0$ units) before a cab is free to respond to the call. At time t_2, the cab completes the response and picks up the customer. This gives a response time of $t_2 - t_1$ units. Finally, the cab delivers the customer at time t_3, for a transit time of $t_3 - t_2$ units.

From a queuing theory perspective, "waiting time" is identical to "response delay," since, technically, the cab begins "service" as soon as the response is initiated. "Service time," in this case, is the sum of "response time" and "transit time." From the customer's (and taxi dispatcher's) perspective, however, service does not begin until the customer is up. "Waiting time," therefore, is the sum of "response delay" and "response time." This is a major reason why queuing theory cannot effectively address the taxi problem. "Waiting time" simply has a different mathematical meaning.

To illustrate the interactions among these response and service variables and to focus on the major issue addressed in this paper, the causal model shown in Figure 4 will be employed. A causal model is interpreted by looking at a relationship between two variables and answering the question, if variable A increases what is the response in Variable B? A positive reaction is indicated by a plus (+) sign and a negative reaction by a negative (-) sign. For example, as shown in the figure, as cab speed increases response time decreases. The relationships form a series of information feedback loops that provide insight into system behavior.

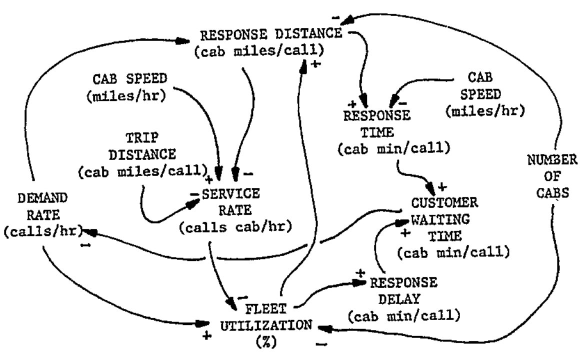

Figure 4: Causal Model Structure

The mathematical structure of the causal model is given in the following equation system:

Waiting Time = Response Time + Response Delay

Response Delay = f(Utilization)

Response Time = Response Distance /Cab Speed

Response Distance = f(Demand, Cabs, Utilization)

Utilization = Demand/Cabs*Service Rate

= Demand/[cabs*[1/((Trip Distance + Response Distance)/Cab Speed)]]

From this it can be seen that as response distance approaches zero; utilization approaches: (demand* trip distance/cabs*cab speed) and service rate approaches: (1/trip time) where trip time equals trip distance/cab speed.

Some of the elements shown are directly under management's control (number of taxis), some cannot be controlled as they are random processes (demand), and some result from the strategies chosen to manage the system (service rate or waiting time). The dispatching strategy, for example, will determine response distance through how a cab is positioned when it is free. Given this, experimentation will focus on the system's behavior resulting from various control policies, differing demand patterns and number of taxis. The impact of changes in demand and in the number of cabs is the focus.

3. EXPERIMENTATION AND CONCLUSIONS

System behavior was investigated by performing a series of experiments with the model. The performance measures used were the "average service rate" and "fleet utilization." It is reasonable to use these measures, because, as shown in Figure 4, they ultimately determine the customer's waiting time. The policy variable was the number of taxis placed in service under various demand patterns. The demand and number of cabs directly impact the performance measures (see Figure 4). While a large number of system operating strategies could be tested with the model, investigation in this research is limited to three representative dispatching strategies.

The first policy involves using only free cabs for allocation to an incoming call with the cab that has been free the longest dispatched (FIFO). Cabs remain at the drop off location. The second policy involves using only free cabs for allocation to an incoming call with the closest cab to the caller dispatched. Cabs remain at the drop off location. The third policy is the same as number two with the exception that all cabs are eligible for selection not just those that are free. In all cases, the customer selection rule is the first customer to call is the first served (FIFO). The candidate policies were chosen to demonstrate the types of interactions prevalent among the variables discussed. It is obvious that extensive experimentation would be required to determine the best operating policy.

Given the structure shown in Figure 4, several expectations about system behavior may be formed. First, for a demand increase, one would expect the response distance to fall, the service rate to rise, and utilization to increase. But there also are expected indirect effects. The rise in service rate would cause utilization to increase which, in turn, would cause the response delay to increase and the response distance to increase. This means that the initial response distance decrease would be tempered by the rise in utilization. The ultimate effect on customer waiting time is very difficult to predict. Second, for a cab increase, the response distance would fall causing a rise in service rate and the fleet utilization would fall. As before, however, there are through the loops, secondary effects, that temper these behaviors. These are most important because the secondary effects on both response time and response delay would lead to the expectations of a significant reduction in customer waiting time. This would lead to the conclusion that waiting time is relatively insensitive to changes in demand but highly sensitive to changes in the number of cabs. (Bailey and Clark, 1985).

The outputs of the model for the three policies are shown in Figures 5 through 10 which will be used to discuss the actual behavior given the suggested expectations. For each policy, either fleet utilization or the average service rate is plotted against a variable demand rate. Each line in the graph represents a response for a given number of cabs. Recall that service rate can be used as an indicator of waiting time with a higher service rate yielding a lower waiting time.

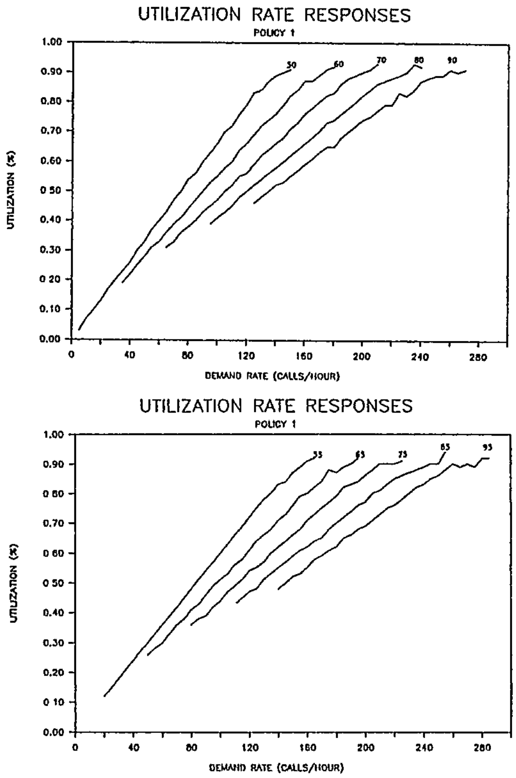

Figure 5: Utilization Rate Response for Policy 1

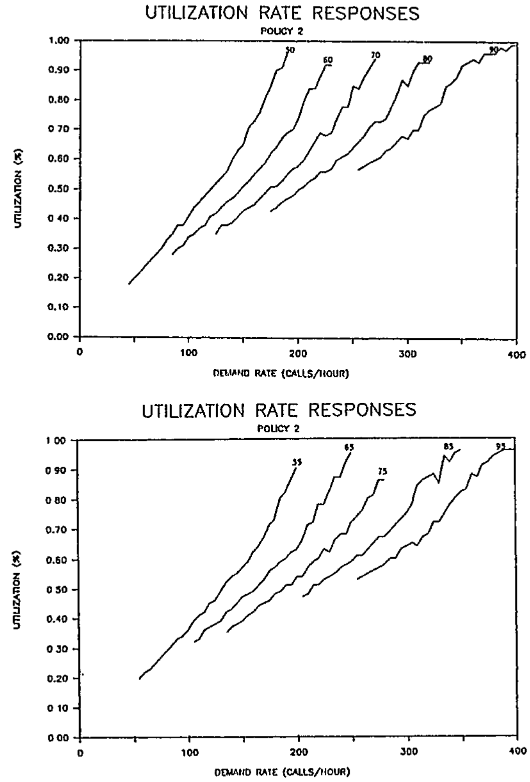

Figure 6: Utilization Rate Responses for Policy 2

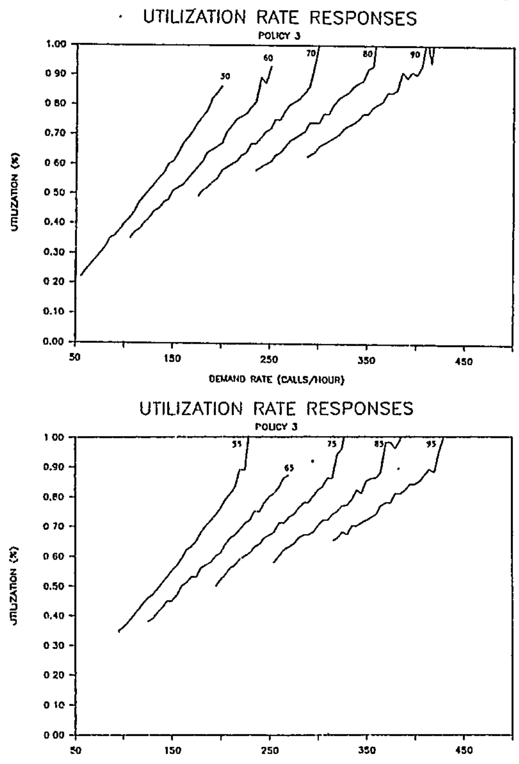

Figure 7: Utilization Rate Response for Policy 3

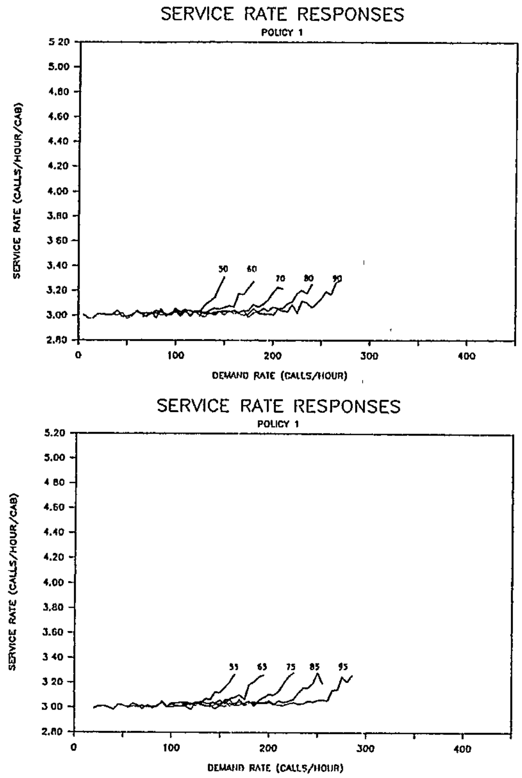

Figure 8: Service Rate Responses for Policy 1

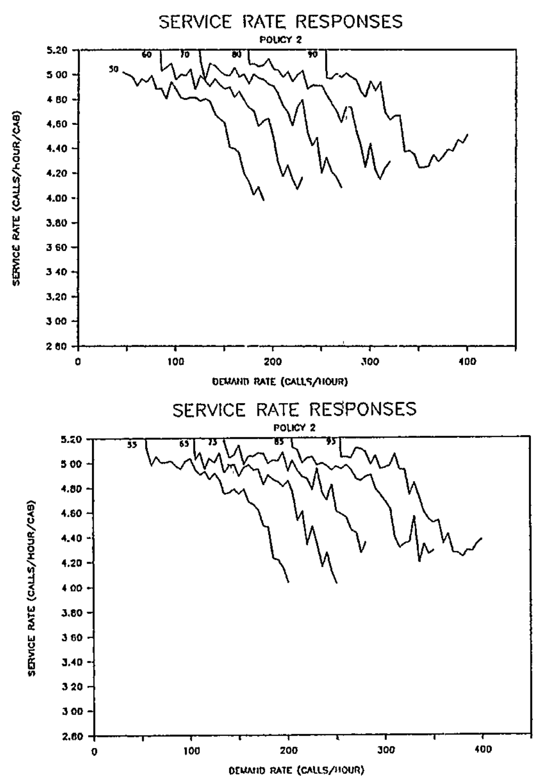

Figure 9: Service Rate Responses for Policy 2

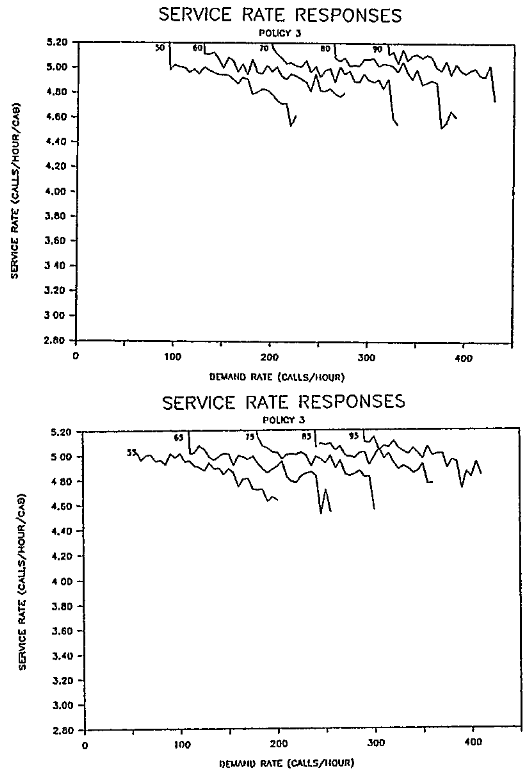

Figure 10: Service Rate Responses for Policy 3

The service rate versus demand plots for policies two and three show that the effect of a demand increase for any constant number of cabs causes a decrease in service rate not the expected increase. This means that the indirect effects that increased utilization (see the utilization versus demand plots) have on response distance more than offset any decrease in response distance resulting from an increase in demand. This is not the case for policy one because at high utilization rates, cabs effectively are no longer free, so the nearest cab for a given call responds.

This means that to correctly predict the number of cabs to add when demand increases cause utilization to approach 100 percent, the relationships in the feedback loops among the variables must be empirically derived. An increase in demand does not significantly increase customer waiting time until fleet utilization approaches 100%. The company should not simply add cabs as demand begins to increase but should add them only as utilization approaches 100%.

The behaviors discussed in the model have been used to illustrate the counter intuitive behaviors that result from complex system interactions. The direction and magnitude of reactions to management initiatives are difficult to measure and precisely control. A system model certainly amplifies a manager's experience and intuition and sharpens judgement in policy analysis.

REFERENCES

Baker, E. (1981). An Algorithm for Vehicle Routing with Time Window Constraints. Management Science Department Report. Coral Gables, FL: University of Miami.

Bailey, W. A., Jr. and T. D. Clark, Jr. (1987). Taxi Management and Route Control: A Systems Study and Simulation Experiment. Working Paper 87.1.2, Center for Information Systems Research, College of Business, The Florida State University, Tallahassee, FL.

Bodin, L. D., B. Golden, A. Assad, and M. Ball. (1983). Routing and Scheduling of Vehicles and Crews. Computers and Operations Research, 10:2, 63-211.

Bodin, L. D. and T. R. Sexton. (1982). The Multi-Vehicle Subscriber Dial-A-Ride Problem. Working Paper No. 82-005. College Park, MD: University of Maryland, Department of Management Science and Statistics.

Boxma, O. J., J. W. Cohen and N. Huffels. (1979). Approximations of the Mean Waiting Time in an M/G/s Queuing System. Operations Research, 27, 1115-1127.

Daganzo, C. F. (1978). An Approximate Analytic Model of Many-to-Many Demand Responsive Transportation Systems. Transportation Research, 12, 325-333.

Gerrard, M. J. (1974). Comparison of Taxi and Dial-A-Bus Services. Transportation Science, 8:2, 85-101.

Gross, D. and C. M. Harris. (1985). Fundamentals of Queuing Theory. (Second Edition). New York: John Wiley and Sons.

Hendrickson, C. T. (1978). Approximate Analytic Performance Models of Integrated Transit System Components. Ph.D. Thesis, Department of Civil Engineering, M.I.T., Cambridge, MA.

Hogue, J. T. and H. J. Watson. (1983). Management's Role in the Approval of Administration of Decision Support Systems. MIS Quarterly, 7:2, 15-26.

Hokstad, P. (1978). Approximations for the M/G/m Queue. Operations Research, 26, 510-523.

Jaw, J. J. (1984). Heuristic Algorithms for Multi-Vehicle Advance- Request Dial-A-Ride Problems. Ph.D. Thesis. Cambridge, MA: Massachusetts Institute of Technology, Department of Aeronautics and Astronautics.

Jaw, J. J., A. R. Odoni, H. N. Psaraftis, and N.H.M. Wilson. (1986). A Heuristic Algorithm for the Multi-Vehicle Advance Request Dial-A-Ride Problem with Time Windows. Transportation Research, 20B:3, 243-257.

Kimura, T. (1983). Diffusion Approximation for the M/G/m Queue. Operations Research, 31, 304-321.

Kirby, R. F., K. V. Bhatt, M. A. Kemp, R. G. McGilliway, and M. Wohl. (1974). Paratransit: Neglected Options for Urban Mobility. Washington, D.C.: Urban Institute.

McLeod, M. G. (1972). The Operations and Performance of a Taxi Fleet. Master's Thesis. Cambridge, MA: Massachusetts Institute of Technology, Department of Civil Engineering.

Meyer, R. F. and H. B. Wolfe. (1961). The Organization and Operation of a Taxi Fleet. Naval Research Logistics Quarterly, 8:2, 137-150.

Pritsker, A.A.B (1986). Introduction to Simulation and SLAM II. (Third Edition). New York: Halsted Press.

Psaraftis, H. N. (1980). A Dynamic Programming Solution to the Single Vehicle Many-to-Many Immediate Request Dial-A-Ride Problem. Transportation Science, 14, 130-154.

Sexton, T. R. (1979). The Single Vehicle Many-to-Many Routing and Scheduling Problem. Ph.D. Thesis. Stony Brook, NY: State University of New York.

Sexton, T. R. and L. D. Bodin. (1980). The Single Vehicle Many-to-Many Routing and Scheduling Problem. with Desired Delivery Times. Working Paper No. 80-014. College Park, MD: University of Maryland, Department of Management Science and Statistics.

Simpson, A. U. and T. F. Carberry. (1978). Estimating the Effectiveness of Shared-Ride Taxi. In Proceedings of the Conference on Taxis as Public Transit. G. J. Fielding and R. Teal, eds. Irvine, CA: University of California, Institute of Transportation Studies and Social Sciences, 133-164.

Takahaski, Y. (1977). An Approximation Formula for the Mean Waiting Time of an M/G/c Queue. Journal of the Operational Research Soceity of Japan, 20, 150-163.

Tijms, H. C., M. H. van Hoorn, and A. Federgruen. (1981). Approximations for the Steady-State Probabilities in the M/G/c Queue. Advances in Applied Probability, 13, 186-206.

Wilson, N.H.M. (1967). CARS - Computer Aided Routing System. Cambridge, MA: Massachusetts Institute of Technology, Department of Civil Engineering, Research Report R67-12.

Wilson, N.H.M. and N. J. Colvin. (1979). Information and Control Systems for Paratransit Services. In Urban Transport Service Innovations. Washington, D.C.: Transportation Research Board Special Report #184, 95-102.

Wilson, N.H.M. and N. J. Colvin. (1977). Computer Control of the Rochester Dial-A-Ride System. Cambridge, MA: Massachusetts Institute of Technology, Center for Transportation Studies, Report 77-22, (DOT UMTA-MA-06-0071).

Wilson, N.H.M. and B. T. Higgonet. (1973). General Purpose Computer Dispatching System. In Demand Responsive Transportation Systems. Washington, D.C.: Transportation Research Board Special Report #136, 102-104.

Wilson, N.H.M. and E. Miller. (1977). Advanced Dial-A-Ride Algorithms Research Project. Phase II: Interim Report. Cambridge, MA: Massachusetts Institute of Technology, Department of Civil Engineering, Research Report R77-31, (DOT UMTA-MA-06-0071).

Wilson, N.H.M., J. M. Sussman, B. T. Higgonet, and L. A. Goodman. (1970). Simulation of a Computer-Aided Routing System (CARS). In Mass Transportation. Washington, D.C.: Transportation Research Board Record #318, 66-76.

Wilson, N.H.M., J. M. Sussman, H. K. Wong, and B. T. Higgonet. (1971). Scheduling Algorithms for a Dial-A-Ride System. Cambridge, MA: Massachusetts Institute of Technology, Urban Systems Laboratory, Report USL TR-70-13.

Wilson, N.H.M., R. H. Weissberg, and J. Hauser. (1976). Advanced Dial-A-Ride Algorithms Research Project. Final Report. Cambridge, MA: Massachusetts Institute of Technology, Department of Civil Engineering, Research Report R76-20 (DOT UMTA-MA-11-0024).

Yao, D. D. (1985). Refining the Diffusion Approximations for the M/G/m Queue. Operations Research, 33, 1266-1277.

AUTHORS' BIOGRAPHIES

WILLIAM A. BAILEY, JR. is an Assistant Professor of Business Administration at The University of North Carolina at Charlotte. He is a member of AEM, TIMS, DSI, and SIM. His research interests include the application of computer simulation to decision support systems and complex transportation systems. He has previously published in WSC.

William A. Bailey, Jr.
Department of Business Administration
College of Business
University of North Carolina at Charlotte
Charlotte, North Carolina 28223

THOMAS D. CLARK, JR. is an Associate Professor of Information and Management Sciences and director of the Center for Information Systems Research at The Florida State University. He has been active in the Winter Simulation Conference for ten years as an author, session chairman and track chairman. He has taught a variety of computer simulation courses at the graduate level and has supervised over sixty masters and doctoral theses and dissertation that have employed simulation analysis. He had extensive experience in logistics management before joining a university faculty and has published a variety of papers using simulation in study of logistics and operations management systems. His current research interests include investigation of a methodology to value management information, the structure of multi-echelon service operations and the nature of decision support system evolution. He is a member of ACM, ORSA, TIMS, SIM, SCS, DSI and SGSR.

Thomas D. Clark, Jr.
Department of Information & Management Sciences
College of Business
The Florida State University
Tallahassee, Florida 32306-1042