Proceedings of the 1986 Winter Simulation Conference
J. Wilson, J. Henriksen, S. Roberts (eds.)

TRAFFIC SIGNAL TIMING AT ISOLATED INTERSECTIONS USING SIMULATION OPTIMIZATION

Anthony A. Saka, G. Anandalingam and Nicholas J. Garber
School of Engineering and Applied Science
University of Virginia
Charlottesville, VA 22901, U.S.A.

ABSTRACT

Two innovative stochastic traffic signal optimization techniques for isolated intersections are discussed. The objective is to determine the optimum cycle and green phase lengths for signalized isolated traffic intersections. Determination of optimum cycle and green phase lengths is based on minimization of the total average delay at the intersection for a given period of observation. Traffic signal timing is formulated as a stochastic inventory problem, which is then solved by a combination of simulation and dynamic programming. The suitability of the optimization techniques for undersaturated and oversaturated flow conditions; and steady state and nonstationary queue conditions are discussed. The advantages of the techniques over most widely used signal optimization methods, and the application of the simulation optimization method in state dependent server-vacation signal timing are also discussed.

1. BACKGROUND

For the past two decades, no significant changes had occurred in the methods used in timing traffic signal at isolated intersections. Most existing models are based on simplified assumptions which give a myopic view of the problem at signalized intersections. Widely used models including the Webster and Reeder models are deterministic, which assume regular traffic arrival and service rate (Webster, 1961). It is known that traffic arrival and service time at signalized intersections are usually randomly distributed. In this light, by assuming regular arrival and service time, errors are often made in determining the optimum cycle and phase lengths of the intersection.

We observed through simulation analysis that in traffic signal timing, particularly as flow approaches saturation, inaccurate estimation of service time can have devastating effects on queue formation and delay at the intersection. For example, in certain flow conditions for a given signal phase, underestimation of the effective green time by one second can increase its average delay by as much as 40 seconds or even more.

Another problem associated with most widely used signal optimization models is that their use is only valid in undersaturated flow conditions. Recently, Cronje' has developed an optimization model for isolated fixed-time signalized intersection that can be used for both undersaturated and oversaturated flow conditions. However, the validity of Cronje's model has only been tested for the M/D/1 queue system i.e., Poisson arrivals and constant service. It is doubtful if the accuracy of the model will be maintained in other arrival and service characteristics.

In this paper, we relax the simplifying assumptions in previous studies and use a simulation approach to (i) estimate optimum signal cycle times for different traffic flow conditions, and (ii) show that assuming regular arrival and service patterns in optimizing traffic signal timing can be very erroneous. The methodology used in this study can be extended to the analysis of protocols in computer and communications systems.

In the following sections, we will discuss the use of alternative fixed and actuated traffic signal timing techniques for isolated signalized intersections that will attempt to eliminate and/or minimize some of the aforementioned problems associated with widely used signal optimization methods.

2. SCOPE/LIMITATION

The traffic signal timing problem presented here is based on the following situations:

1) The intersection has a single server, traffic signal, which provides service to a single signal phase at a time. Two cases will be considered. The first is where the server renders service on a rotational basis. Here, the server is programmed to give fixed service time (green time) to the individual signal phases. After serving a given signal phase for a predetermined service time, the server switches service to another signal phase and continues the process on rotational basis. The second case is where the server provides service to a given signal phase only when its demand satisfies certain stipulated criteria. In other words, service is state-dependent. Here, the server does not have to provide service on rotational basis. Also, the amount of service time to be provided for a given signal phase depends on the state of its demand (queue) and the states of the demand in other signal phases.

2) It is assumed that no time is lost as the server switches service from one phase to another signal phase. In other words, the time it takes the traffic signal to change from one color, say red, to another color, say green, is negligible.

3) The server operates without interruption.

4) The server does not have to satisfy all the demand in a given signal phase before switching to another. The objective is to provide service in such a way that the average delay at the intersection is reasonably minimized.

5) The queue systems considered in the study are limited to the following: i) M/M/1; ii) M/D/1; iii) $M/E_k/1$; iv) $E_m/M/1$; and v) $E_m/E_k/1$. The FIFO queue discipline is used.

3. DATA REQUIREMENT

Three categories of data are required to undertake the signal optimization analysis discussed here. The first category of data is the vehicular mix at the intersection. Vehicles may be classified in the following groups: i) passenger cars, ii) trucks, iii) buses, and iv) motorcyles. Traffic signal timing which does not require high accuracy may assume all vehicles at the intersection as passenger cars.

The second category of data is the arrival headways, and the individual service time for the above four classes of vehicles. Service time is defined as the time used to discharge the individual vehicles from the intersection during the time the traffic light stays green. This should not be confused with the total service time of a given signal phase, which is the effective green time or green phase length. Sometimes, traffic engineers also collect data on lost service time, i.e. the time wasted before the first vehicle in the queue starts to move as the traffic light changes from red to green. Since lost service time is usually independent and randomly distributed, it is herein not considered separately, but instead added to the service time of the first vehicle in the queue.

The third category of data is the average arrival and service rate of pedestrians at the intersection. This information is needed to determine the all-red phase length for pedestrian crossing. It is customarily to use the data on the approach with the largest pedestrian traffic for the signal timing.

4. DATA ANALYSIS

In this paper, traffic arrival and service times at a given intersection are considered as independent random variables, with known distributions. In this regard, the data analysis involves determination of appropriate arrival and service distributions for the individual approaches at the intersection. Typically, 95% confidence interval is considered adequate in selecting the appropriate distributions to fit the data.

Because of the random nature of traffic arrival, the Poisson distribution usually makes a good fit. the memoryless nature of the Exponential distribution has made it widely accepted by researchers in fitting randomly distributed service times, such as those at signalized intersections.

The Erlang distribution, which is halfway between complete randomness and degeneracy is perhaps the most suitable distribution for traffic studies at signalized intersections. The probability density function of the Erlang distribution is given as (Hillier and Lieberman, 1980):

$$f(t) = \frac{(\alpha k)^k t^{k-1} e^{-k\alpha t}}{(k-1)!}, \text{ for } t \geq o \qquad (1)$$

where

$1/\alpha$ = mean of the distribution

$\frac{1}{\sqrt{k}} \frac{1}{\alpha}$ = standard deviation of the distribution

α = "scale factor" and is a positive number

k = "shape parameter" which characterizes the degree of variability of the distribution, and is a positive integer.

t = arrival or service time variable

The importance of the Erlang distribution can be appreciated as it becomes an Exponential distribution when the value of k = 1 and a constant (degenerate) distribution when the value of $k = \infty$.

It is suggested that the Erlang distribution be considered before any other distribution in selecting the appropriate distributions to fit traffic arrival and service time data for signalized intersections.

The following information is required from the data analysis:

i) Estimation of the parameters associated with the distributions considered. For example, in the case of Erlang distribution, parameters α and k are to be estimated.

ii) Selection of appropriate arrival and service time distributions for the individual approaches at the intersection, and the level of significance of the error associated with each distribution.

iii) The length of time to be used in running the simulation in the case of nonstationary queue conditions. Since traffic arrival at signalized intersections is a dynamic process, changing from one state to another as we move from one time stage to another, it is suggested that signal optimization be undertaken in stages. This is particularly necessary in urban areas where significant variations occur between peak and off peak hour traffic flow.

iv) The appropriate all-red phase length for pedestrian crossing.

v) The average service time for each of the aforementioned four classes of vehicles, and the percentage of the total traffic each class of vehicle represents for a given approach.

5. MODELLING

Traffic signal timing in this paper is modelled as a stochastic inventory problem, with no set up cost and where unsatisfied demand is backlogged. In other words, demand which is not satisfied stays in queue for the next available service time.

Consider a four phase signalized intersection like the one shown in Figure 1. The queues associated with the individual signal phases are herein considered as our commodities on which inventory is to be made. Since the signalized intersection is a case of a single server, green time for a given signal phase is red time for other signal phases.

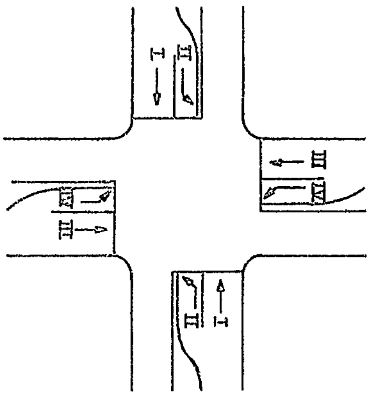

Figure 1: A Four-Phase Signalized Intersection

For example, in Figure 1 when phase I has green light, phases II, III, and IV will have red light. In this regard, by serving the queue in phase I, we accumulate more queues in phases II, III, and IV. Since this is an inventory problem, delay associated with the queue in each signal phase is considered as the holding cost.

The holding cost associated with a given green signal phase length, t_j can be estimated as:

$$D_j = \sum_{i=1}^{N-1} A_i \int_o^{tj} \xi \, \Psi_{D_i}(\xi)\, d\xi + Aj \ \mathrm{Max}\left\{\left[\int_o^{tj} \xi \, \Psi_{D_j}(\xi)\, d\xi - \left(\sum_{a=1}^{M} \int_o^{tj} \alpha \, \Psi_{D_j}(\phi(\alpha_a))\, d\alpha\right)\right], o\right\} \tag{2}$$

where

- D_j = Expected delay (holding cost) associated with signal phase, j.
- A_i = Multiplicative factor which converts queue to delay in phase, i.
- t_j = Service time (effective green time) for signal phase, j.
- $\psi_D(\cdot)$ = Probability density function of arrival/-service time for individual vehicles.
- ϕ = Discrete probability function that characterizes vehicular mix.
- ξ = Random arrival.
- α = Random service.
- M = Total number of vehicular classes.
- N = Total number of signal phases.

The two components of the right hand side of the above delay expression are the delays in red signal phases and the green phase, respectively if service time, t_j is assigned to green phase, j.

$$\text{Let } D_T = \sum_{j=1}^{N} D_j + \sum_{i=0}^{N} A_i \int_o^{R} \xi \psi_{D_i}(\xi) d\xi \tag{3}$$

where

R = All-red phase length.

The optimum inventory policy is the selection of service time, t_j^*, for each signal phase, j, such that:

$$D^* = \mathrm{Min}\left\{ D_T \right\} / N \tag{4}$$

where

- D^* = Minimized average delay per vehicle at the intersection
- t_j^* = Optimum service time for signal phase, j

It should be noted that the multiplicative factor, (A) which converts queue to delay is actually dynamic, and depends on the interarrival times, which we have earlier considered to be independent and random. Therefore, accurate determination of (A), given the queue size will involve determination of interarrival times of the individual vehicles in the queue. This is difficult to do analytically, particularly in large problems with highly fluctuating traffic flow.

The delay expressions derived for the inventory problem are used only in the dynamic programming model (described in section 6.2). There, we approximated (A) as the average interarrival time.

Because simulation clocks the time each arriving entity spends in the system before being served; in the simulation optimization which we discussed in section 6.1, we directly obtained the delays for the individual signal phases from the simulation output.

6. OPTIMIZATION

The optimization policy considered here is minimization of total expected delay at the intersection, given a range of experimental cycle lengths. Two optimization methods are considered: 1) Simulation and 2) Dynamic Programming. We provide empirical results for the simulation case. The coding of the dynamic programming algorithm has not been completed as yet although we briefly discuss how it can be used in determining the optimum cycle and phase lengths.

6.1 Simulation Optimization

The microcomputer version of SIMAN is used for the simulation modelling (Pegden, 1985). For each signal phase of the intersection, we developed a simulation model to obtain the expected queues and delays. The models account for: a) interarrival and service distributions and the mean rates; b) distribution of different classes of vehicles being served at the intersection; c) experimental cycle and phase lengths; and d) the period of time during which observations are made. The study results presented in this paper are for two hours observation period. However, the performance of the system at any given time period can be traced during the simulation. This enables the analyst to determine information such as the delay each arriving entity experienced while waiting in the system to be served.

Optimum cycle and green phase lengths for the aforementioned inventory problem are determined by trial and error procedure, using different values of experimental cycle lengths. The range of experimental cycle lengths to be considered in the optimization process depends on the experience of the engineer or analyst.

In practice, cycle lengths which exceed 120 seconds are usually considered undesirable. For a two-phase signalized intersection, the minimum desirable cycle length is usually 30 seconds (Michalopoulos and Plum, 1983).

Advantages of Simulation Optimization. The use of simulation for optimizing isolated intersection traffic signal timing has several advantages over most widely used methods. First, simulation does not use empirically derived models to estimate variables such as queues and delays which are used to determine the optimum cycle and green phase lengths of the intersection. In this regard, it is considered by many as the most accurate method.

Second, because packages that are available at present (for example, SIMAN) are simple to use, it is easier to solve complex problems using simulation rather than mathematical programming formulations which require complex software engineering.

Third, most widely used traffic signal optimization models are limited to certain traffic condition(s); i.e. undersaturated flow, steady state queues and delays, regular arrival and service time, etc. Simulation optimization is not limited by any of these conditions. However, it must be stressed that simulation optimization is not cost effective if used for steady state queue/delay conditions. This is primarily due to the fact that it takes long period of running time to achieve a steady state queue/delay condition, particularly as flow approaches saturation.

Fourth, most widely used optimization methods are limited to either fixed or actuated-signal timing. Because the minimizing queues and delays for the individual signal phases can be obtained from the simulation output, simulation optimization method can conveniently be used for both fixed and actuated signal timing. Its application in a special type of actuated signal timing herein referred to as "state dependent server-vacation signal timing" is discussed in the next subsection (Harris and Marchal, 1986).

State Dependent Signal Timing. An important aspect of simulation optimization is its application in state dependent server-vacation signal timing, where the server goes on vacation when the state of the queue in a given approach satisfies certain stipulated criteria. Those approaches where the server is on vacation are given red light and those which receive service from the server are given green light.

A major advantage of the state dependent server-vacation approach is that service times are more intensively used than most type of actuated-signal timing methods. There is less wastage of service (green) time because service is rendered only when there is adequate demand for it.

In the case of most semi-actuated signal timing, the server is obligated to render the predetermined minimum service regardless of whether there is adequate demand or not. The state dependent server-vacation approach, therefore, eliminates or minimizes the uncertainties associated with fluctuations in traffic arrivals at signalized intersections.

Programming the Server. The flow chart in Figure 2 illustrates how the server in state dependent server-vacation signal timing may be programmed.

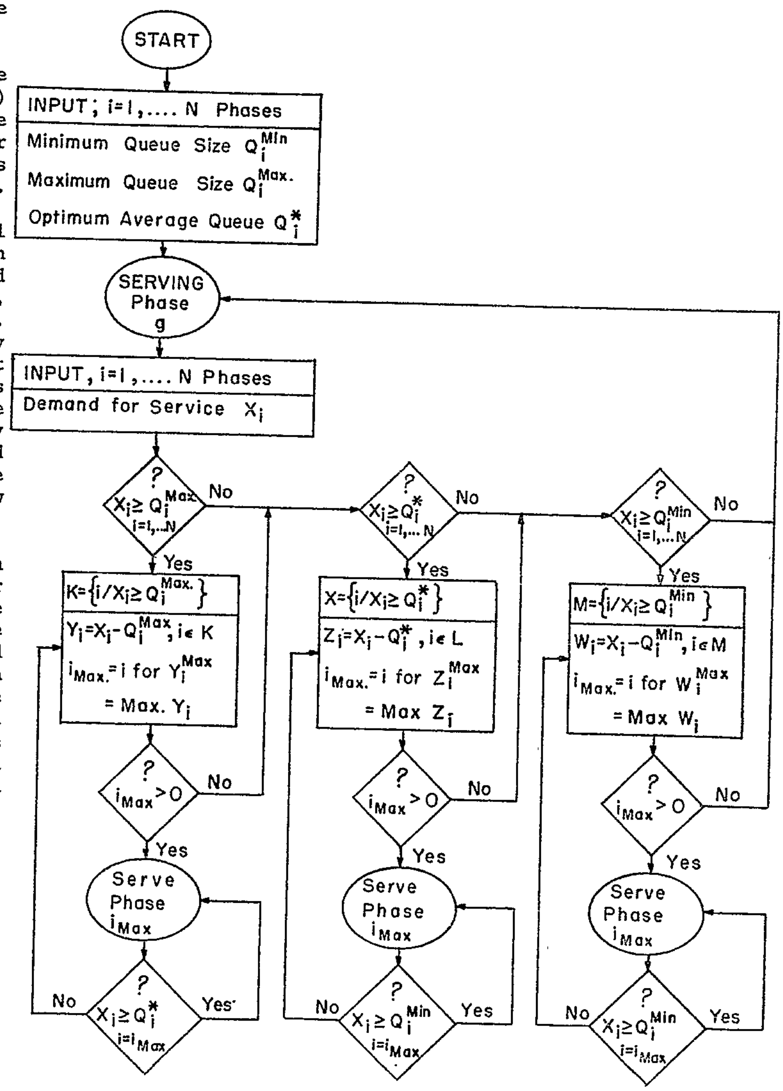

Figure 2: A Simplified State Dependent-Server-Vacation Signal Timing

In Figure 3, it can be observed that depending on the state of queues in other phases, the server can go on vacation at any point in the vacation zone and can resume service at any point in the service zone. However, the server is not allowed to let the queue in any approach to exceed its maximum tolerable queue size. In this light, the server can occasionally go on premature vacation under certain flow condition in order to prevent queue at a given approach from exceeding its maximum tolerable queue size.

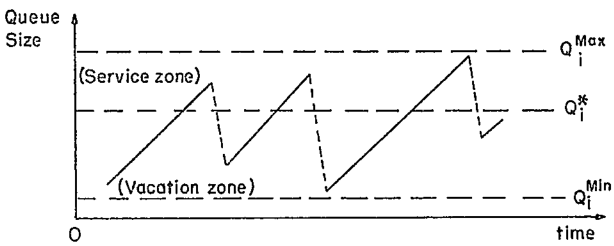

Figure 3: State Dependent Server-Vacation Queue Schematic

6.2 Optimization by Dynamic Programming

The dynamic program used to determine the optimum policy of the previously discussed inventory problem is:

$$\text{Minimize} \quad \sum_{j=1}^{N} D_j\ (t_j) \ + \ \sum_{i=1}^{N} D_i\ (R) \qquad (6)$$

$$\text{Subject to} \quad R \ + \sum_{j=1}^{N} t_j \ = \ C$$

$$t_j > 0$$

$$R \geq 0$$

where

$D_j(t_j)$ = Expected delay associated with green phase, j, having green time, t_j.
$D_i(R)$ = Expected delay in red phase, i, due to All-red phase length, R.
N = Total number signal phases
C = Cycle length of the intersection.

The individual signal phases are used as the stage elements, and the varying amount of effective green time (service time) available for allocation are used as the state elements of the Dynamic Programming.

The above dynamic program is solved using the well known backward recursive computation method (Hillier and Lieberman, 1980). Here, too, the optimum cycle length of the intersection is determined by experimenting with several values.

7. STUDY RESULTS

In this section we provide results based on simulation on the effect of different arrival and service time distributions on the optimum cycle length of a signalized traffic intersection keeping hourly traffic volume and average service rate constant.

We consider a four-phase signal intersection with hourly volume of 600, 450, 350, and 300 vehicles, respectively. We assume an average service rate of 0.5 vehicles/second and an all-red phase length of 10 seconds. It is shown in Tables 1, 2 and 3 that the optimum cycle length of the intersection is different for M/M/1, M/D/1, E_2/M/1, and D/D/1 queue systems.

Table 1: Signal Optimization Output for M/M/1

Experimental Cycle Lengths	Average Delay Per Vehicle at Intersection
120 seconds	72.5 seconds
110 seconds	72.7 seconds
100 seconds	71.4 seconds*
90 seconds	73.6 seconds

Table 2: Signal Optimization Output for M/D/1

Experimental Cycle Lengths	Average Delay Per Vehicle at Intersection
120 seconds	77.9 seconds*
110 seconds	82.1 seconds
100 seconds	94.4 seconds
90 seconds	95.5 seconds

Table 3: Signal Optimization Output for E_2/M/1

Experimental Cycle Lengths	Average Delay Per Vehicle at Intersection
120 seconds	27.4 seconds
110 seconds	25.9 seconds
100 seconds	23.7 seconds
90 seconds	23.1 seconds*

Table 4: Signal Optimization Output for D/D/1

Experimental Cycle Lengths	Average Delay Per Vehicle at Intersection
120 seconds	72.09 seconds
110 seconds	69.70 seconds*
100 seconds	86.85 seconds
90 seconds	87.32 seconds

For the same data, the optimum cycle length is 100 seconds if the M/M/1 queue system is assumed, 120 seconds if the M/D/1 queue system is assumed, 90 seconds if E_2/M/1 queue system is assumed and 110 seconds if D/D/1 queue system is assumed. It can be deduced from the above results that by assuming wrong arrival and service patterns, the engineer or analyst is likely to use the wrong cycle length, which may even cause more delays at the intersection.

For example, by wrongly assuming a Poisson arrival as regular arrival the optimum cycle length changes from 120 seconds to 110 seconds. This increases the average delay per vehicle by over four seconds or the total vehicular delay at the intersection by about two hours per hour.

It is also important to note that since most widely used signal optimization models are based on steady state queue conditions, they tend to overestimate the optimum cycle length, particularly as flow approaches saturation. For example, the Webster model determined the optimum cycle length of the above problem to be 360 seconds.

8. VALIDATION OF RESULTS

We validate our results by using different streams of random numbers in running the simulation. Although the optimum solutions remain unchanged for the different streams, the queue sizes and delays for the individual signal phases vary slightly from stream to stream.

In state dependent server-vacation signal timing where criteria are set on the minimum, average and maximum tolerable queue sizes for the individual signal phases, it is suggested that different streams of random numbers be used in running the simulation. The overall averages of the optimum minimum, average and maximum queue sizes obtained by considering different streams should be used in programming the state dependent server.

There are ten different streams of random numbers available in SIMAN. The number of streams to be used in the simulation will depend on the analyst. However, based on the law of large numbers the validity of the study results are expected to be increased by increasing the number of streams to be used in running the simulation.

9. CONCLUSIONS

It has been shown in this paper that assuming regular arrival and service patterns in optimizing traffic signal timing is very erroneous, and may increase the magnitude of the problems at signalized intersections that the engineer tries to overcome.

Also, since traffic arrival at signalized intersections is a Markov process, changing from state to state within a short period of time, the use of optimization models which are based on steady state queue conditions is considered inappropriate, particularly in heavy traffic conditions. When traffic is heavy, it usually takes a long period of time (usually longer than peak hour period) to attain steady state queue conditions.

The best traffic signal optimization model or method is that which can be applied effectively in any queue condition. As far as it is known, no existing method but simulation has such capability.

REFERENCES

Bertksekas, D. B., Dynamic Programming and Stochastic Control, New York: Academic Press, 1976.

Cronje, W. B., "Optimization Model for Isolated Signalized Traffic Intersections," Transportation Research Record 905, TRB, Washington, D.C., 1983.

Harris, C. M. and Marchal, W. G., State Dependent Server-Vacation Queues for Modeling Local Area Networks, Report No. GMU/41501/101, March 1986.

Hillier, F. S. and Lieberman, G. J., Introduction to Operations Research, Oakland, California: Holden-Day, Inc., 1980.

Michalopoulus, P. G. and R. Plum, "Determining Capacity and Selecting Appropriate Type of Control at One-Lane Two-Way Construction Sites," Transportation Research Record 905, TRB, Washington, D. C., 1983.

Pegden, Dennis C., Introduction to SIMAN with Version 3.0 Enhancements, Systems Modeling Corporation, State College, Pennsylvania, July 1985.

Takas, L., "Two Queues Attended by a Single Server," Operations Research 16, 1968.

Webster, F. V., Traffic Signal Settings, Road Research Technical Paper No. 39, British Road Research Lab, 1961.

AUTHORS' BIOGRAPHIES

ANTHONY A. SAKA is a Ph.D. student in Transportation Engineering at the University of Virginia. He received a B.A. in City and Regional Planning from the University of Southwestern Louisiana in 1981, and a M.U.R.P. from the University of Pittsburgh in 1982. His current research interests include Mathematical Programming Applications in Transportation Engineering, Transportation Demand Modeling, and Traffic Accidents Analysis. He is a student member of ITE and APA.

Anthony A. Saka
Department of Civil Engineering
School of Engineering and Applied Science
University of Virginia
Charlottesville, VA 22901, U.S.A.
(804) 295-7934

G. ANANDALINGAM is an Assistant Professor of Systems Engineering at the University of Virginia. He received his B.A. (1975) from Cambridge University, England and his Ph.D. (1981) from Harvard University. He worked as an Energy Systems Analyst at Brookhaven National Laboratories (1981-1984). He is an associate contributing editor of International Abstracts in Operations Research and a member of ORSA and TIMS. HIs current research interests include optimization, multi-objective decision making, environmental systems analysis, bargaining and game theory, and forecasting.

G. Anandalingam
Department of Systems Engineering
School of Engineering and Applied Science
University of Virginia
Charlottesville, VA 22901, U.S.A.
(804) 924-7165

N. J. GARBER obtained the B.Sc. degree in civil engineering from the University of London in 1961 and M.S. and Ph.D. degrees from Carnegie-Mellon University in 1968 and 1970, respectively. He is an Associate Professor at the University of Virginia. Dr. Garber has directed several projects as a consultant to engineering firms, and some of these projects were carried out for international agencies such as the World Bank, European Economic Commission and AID. He is a registered Professional Engineer in Virginia. His current research interests are in traffic engineering and highway safety.

N. J. Garber
Department of Civil Engineering
School of Engineering and Applied Science
University of Virginia
Charlottesville, VA 22901, U.S.A.
(804) 924-6366