Transportation Science

Modeling Vehicular Traffic Flow using M/G/C/C State Dependent Queueing Models

Rajat Jain, J. Macgregor Smith,

To cite this article:
Rajat Jain, J. Macgregor Smith, (1997) Modeling Vehicular Traffic Flow using M/G/C/C State Dependent Queueing Models. Transportation Science 31(4):324-336. http://dx.doi.org/10.1287/trsc.31.4.324

Modeling Vehicular Traffic Flow using M/G/C/C State Dependent Queueing Models

RAJAT JAIN and J. MACGREGOR SMITH

Department of Mechanical and Industrial Engineering, University of Massachusetts, Amherst, Massachusetts 01003

In this paper, M/G/C/C state dependent queueing models are proposed for modeling and analyzing vehicular traffic flows. Congestion aspects of traffic flow are represented by introducing state dependent service rates as a function of number of vehicles on each road link. Analytical models for unidirectional and multisource flows are presented. Finally, queueing models to analytically determine the optimal capacity and performance measures of the road links are incorporated into a series of software programs available from the authors.

The management of roadway systems to improve the quality of vehicular travel has always been one of the most interesting and challenging problems in highway traffic design and management. As the number of vehicles approaches or exceeds the existing highway capacity, the movement of vehicles deteriorates at an exponential rate and, unfortunately, this results in over-saturated highways which are not able to efficiently handle the growing demands of traffic.

Researchers from widely varying disciplines such as transportation engineering, civil engineering, and industrial engineering, have utilized different techniques to study traffic interaction and movement. Some of the analytical techniques that have been used for analyzing traffic flow along roadways include capacity analysis (MOSKOWITZ and NEWMAN 1963, BERRY 1977, TRANSPORTATION RESEARCH BOARD 1985a), shock wave analysis (RICHARDS 1956, FRANKLIN 1961, PIPES 1965, WIRASINGHE 1978), queueing analysis (BELL 1980, KUWAHARA and NEWELL 1987), and simulation modeling (FOX and LEHMAN 1967, WIGAN 1969, ELDOR and MAY 1975, TRANSPORTATION RESEARCH BOARD 1981). We approach the vehicular traffic flow problem from a queueing theory perspective. Queueing analysis in the past has mainly been utilized for evaluating traffic flow using *deterministic* models (MAY and KELLER, 1967), for traffic light synchronization (NEWELL, 1965), for analyzing movement of vehicles at intersections, and other performance measures using traditional queueing models. In this paper, we develop a generalized M/G/C/C state dependent queueing model to analyze the performance of a congested roadway segment. It must be pointed out here, that other queueing models could be used to analyze different problems arising in traffic flow. However, we believe that M/G/C/C state dependent queueing models with finite capacity (C) can represent congestion in traffic flow more appropriately and accurately. This representation captures the features of roadway segments which can accommodate a finite number of vehicles and the deterioration in service time as a function of the number of vehicles occupying the roadway.

Over the years, we have developed state-dependent queueing models primarily to analyze movement of pedestrians within facilities such as schools, hospitals, manufacturing plants and related structures (YUHASKI and SMITH 1989, CHEAH 1990, SMITH 1991, CHEAH and SMITH 1994). Our key focus here is on modeling stochastic traffic flow of vehicles traversing the roadways and on developing congestion models to study this traffic flow. In addition, we will illustrate how these queueing models can be used for determining the extreme values of several design and control parameters of highway systems.

1. PROBLEM DEFINITION

CONGESTION OCCURS MAINLY as a result of the increased number of vehicles (customers) competing

Transportation Science
Vol. 31, No. 4, November 1997

0041-1655/97/3104-0324 $01.25

for the limited space available on the roadway segment. Customers driving from any given origin node to any given destination node tend to choose those road segments that lie on the route with shortest travel time. This behavior results in an under-utilization of certain roadlinks while other links are subjected to more traffic than they are designed to handle, thereby resulting in congestion of such links. In essence, an appropriate model needs to be developed to study the dynamic and stochastic effects of congestion on the stochastic traffic flow. However, before we can determine the congestion effects for the whole road network, we need to discuss how to model congested flow along a one-way segment of a traffic network.

Two essential components which can be used to analyze congestion are:

1. *Decay of the service rate within the roadways as a direct result of increased vehicular traffic demand.* In queueing terminology, service rate is defined as a description of time to complete a service, and of the number of customers whose requirements are satisfied at each service event.
2. *There is a finite amount of available space within the roadway segments.*

Analyzing these components leads to the question of how to model this congestion within the roadway segments. What models have been used in the past: in theory and in practice? How can we "best" capture the congestion effects from a queueing theory perspective?

Assumptions

Before proceeding to analyze congestion effects, we make the following assumptions:

1. The arrival process, or pattern of arrivals into the system, is defined by the probability distribution of the number of customers or units that appear at each of the arrival events. For the sake of argument, we'll examine arrivals as independent Poisson processes with a rate λ. Bursty arrivals or nonstationary ones are not considered in the present paper. (Since only one vehicle can enter a single lane at any one time, we feel that the individual vehicular arrivals is a reasonable assumption. Another technique would be to look at the overall road segment where the concept of *platooned arrivals* could be used to represent the flow (ALFA and NEUTS, 1995).)
2. Increasing density of the vehicles within a roadway link leads to a decay in the service rate of the segment.

For any given road segment, we define the following parameters, which describe the physical aspects of the link under consideration.

System Parameters

1. *Capacity/Number of Service Channels*: The *capacity* of a single road segment is defined to be the maximum number of vehicles that can be accommodated on the road link at any given time. We denote this upper bound on the number of vehicles by C. It should be noted that we use the term "capacity" (which refers to number of service channels and buffer space) in a queueing context and this "capacity" is not the maximum expected flow rate (number of vehicles/time unit) as is defined in traffic science. Furthermore, the space occupied by an individual vehicle on the road segment represents one queueing "server." To clarify further, this "server" starts service as soon as a vehicle joins the segment and carries the "service" (the act of traveling) until the end of the segment is reached.
2. *Types of Road Links/Nodes*: Different types of roadways such as multilane freeways, ramps, etc. represent road links with varying characteristics in terms of capacity, construction, and design.
3. *Service Rates*: The service rate for a given road link is the speed with which the vehicles traverse that node. Since the average travel speed deteriorates with increasing number of vehicles (and vice versa), the service rate is state-dependent and depends on the number of vehicles in the system. We define V_n to be the average travel speed of n vehicles on a road link and A to be the average travel speed of a *lone* occupant ($A = V_1$).
4. *Vehicle Equivalents*: There are several different types of vehicles that utilize the roadways, such as passenger cars, trucks, buses, and recreational vehicles. For sake of simplicity, these different vehicle types are assumed to be identical and are termed *passenger car units*. A conversion format, described in (TRANSPORTATION RESEARCH BOARD, 1985b), can easily be used to convert different vehicle types in terms proportionate with a passenger car.

The modeling of state-dependent queues originated with the work of CONWAY and MAXWELL (1962), who studied the nonlinear effects of increasing traffic on the service rate of a server. This work was later extended to the multiple-server case by HILLIER, CONWAY, and MAXWELL (1964).

In their model, if service rate increases, then

$$\mu_n = n^{\gamma}\mu,$$

where

n = the number of units/customers in the system
μ_n = the mean service rate when there are n customers in the system
$1/\mu$ = the mean "normal" service time—the mean time to service a customer when that customer is the only one in the system
γ = the "pressure coefficient"—a constant that indicates the degree to which the service rate of the system is affected by the system state ($\gamma \geq 0$).

However, this model is not defined for the situation where increasing traffic results in a deterioration of service rates, namely for the case when $\gamma < 0$. In the following section, we are able to capture this *decaying* service rate by developing finite, state-dependent congestion models.

2. CONGESTION MODEL

IN VEHICULAR TRAFFIC FLOW, the vehicle density on the roadways largely effects the speed of the vehicles traversing the roadway link. As the vehicular density increases, the average travel speed of the vehicles deteriorates. Other factors such as number of lanes, vehicle types, percentage of heavy vehicles, paving pattern, topography, etc. also affect traffic flows (Transportation Research Board, 1985b), however we attempt to describe congestion in terms of vehicular density veh/mi-lane.

All roadway systems consist of links which have moving vehicles that interact with each other and with the highway pavement. To better understand congestion within roadway networks, we consider a unidirectional road segment of length (miles), L, width, W, and consisting of N lanes.

The maximum capacity for a given road-link is an integral concept with respect to our congestion modeling. Different empirical studies provide different estimates (ranging from 185–265 veh/mi-lane) for that vehicular density at which flow comes to a halt (or *jam density*). However, the varying conditions/factors under which different studies were undertaken makes it impossible to compare one test result with another. We define this jam density value as a constant k veh/mi-lane. We define the capacity of a roadlink to be the highest integer that is less than k times the length of the link and the number of lanes on the link. Thus, the capacity is expressed as,

$$C = [k * L * N]$$

where

C = Capacity of the Road link (vehicles),
k = Constant determined by jam density of appropriate context (veh/mi-lane),
L = Length (miles) of the road link, and
N = Number of Lanes on the Road link.

The available empirical traffic flow models in the existing literature primarily belong to a family of either *Noncongested (Free) flow* models or *Congested flow* models [refer to EASA and MAY (1980) for details]. These models describe relationships between the vehicular density and mean travel speeds. By carefully approximating and fitting the positions of three representative points among the curves recreated from the formulas provided in MAY (1990) for some of the Free flow models based on EDIE (1961), GREENSHIELDS (1935), DRAKE, SCHOFER, and MAY (1967), and UNDERWOOD (1961), and for some of the Congested flow models based on Drake, Schofer, and May (1967), we were able to develop linear and exponential relationships for the empirical models. Our models for roadway traffic flow are based on the methodology originally developed by Yuhaski and Smith (1989) for unidirectional flow within buildings.

We define,

V_n = Average travel speed for n vehicles within a road link,
V_a = Average travel speed at vehicular density = 20 veh/mi-lane,
V_b = Average travel speed at vehicular density = 140 veh/mi-lane,
A = V_1 = Average travel speed of a lone occupant.
n = Number of vehicles utilizing the road link.
γ, β = Shape and Scale parameters for the exponential model.

As the number of vehicles (n) approaches the road link capacity (C), all vehicular traffic flow comes to a halt. Since $n = C + 1$ is an impossible scenario, we set $V_n = 0$ for all cases where $n \geq C + 1$. Using $A = V_1$ as the travel speed of a single vehicle (same as the posted speed limit) and $V_{C+1} = 0$, we get the linear relationship as,

Linear Model:

$$V_n = \frac{A}{C}(C + 1 - n)$$

Figure 1 shows linear model curves plotted for different values of jam density (k = 185, 205, 210, and 265 veh/mi-lane) where mean speed was considered as 55 mph.

The empirical curves for vehicular traffic flow strongly suggest that an exponential relationship may provide a more accurate approximation for the

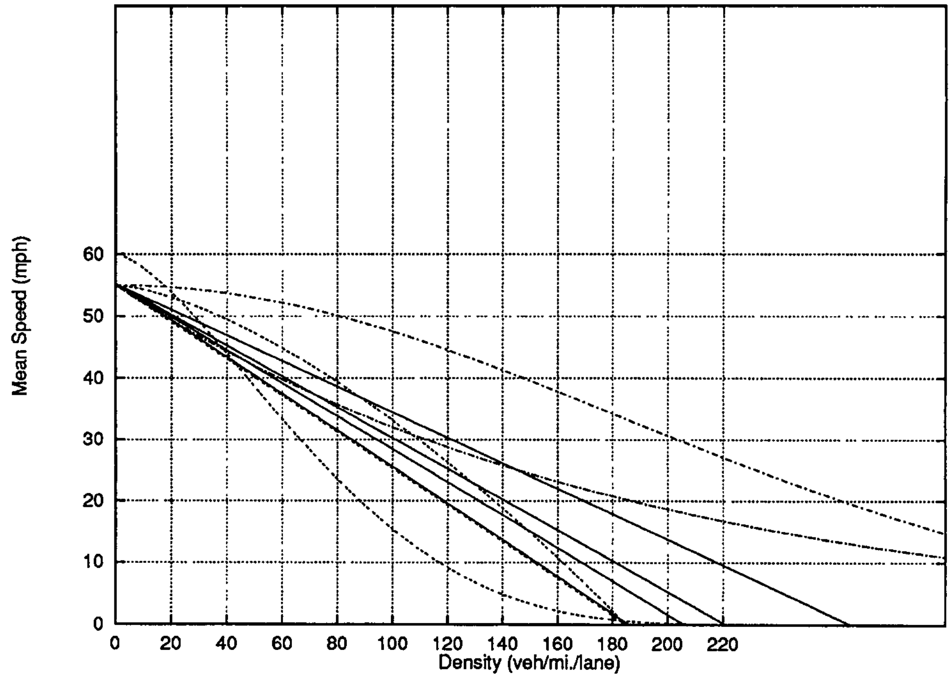

Fig. 1. Linear approximation of empirical traffic flow models at mean speed = 55 mph. MGCC-LIN (a, b, c, d) (———); Underwood (– · – · –); Edie (– – – – –); Greenshields (- - -- - -); HCM (- - - - - -); Drake (- - - - - -).

mean travel speed with variation in vehicular density. Using β and γ as scale and shape parameters, respectively, we developed the following exponential relationship:

Exponential Model:

$$V_n = A \exp\left[-\left(\frac{n-1}{\beta}\right)^{\gamma}\right]$$

where

$$\gamma = \ln\left[\frac{\ln(V_a/A)}{\ln(V_b/A)}\right] \Big/ \ln\left(\frac{a-1}{b-1}\right),$$

$$\beta = \frac{a-1}{[\ln(A/V_a)]^{1/\gamma}} = \frac{b-1}{[\ln(A/V_b)]^{1/\gamma}},$$

and,

$$A = V_1 = 55 \text{ mph} \quad (A = V_1 = 60 \text{ mph})$$

$$V_a = 48.0 \quad (V_a = 50.0)$$

$$V_b = 20.0, \quad (V_b = 16.0)$$

$$a = 20 * L * N, \quad (a = 20 * L * N)$$

$$b = 140 * L * N \quad (b = 140 * L * N).$$

Figure 2 shows curves of this exponential model fitted among the empirical relationship curves for a mean speed of 55 mph.

It should be noted here that other approximation techniques such as weighted nonlinear regression and piece-wise linear approximations could also have been used instead of three-point approximation. However, we believe that, at this stage of research, the impact of any difference between the results obtained from different approximation techniques would be insignificant.

3. ANALYTICAL MODEL: SINGLE LINK

IN THIS SECTION, we discuss the methodology used in modeling a unidirectional single road segment as a queueing system. The road segment under consideration has N lanes, is of length L, and has a capacity $C = [k\ L\ N]$.

Suppose that the vehicles arrive to the road link as a Poisson process with some rate, λ. We assume that the service times of the vehicles follow a general distribution G and that the service rate, $f(n)$, is dependent on the number of vehicles (n) in the system. This state dependent service rate steps "up" or "down" to $f(n+1)$ or $f(n-1)$ when an arrival or a departure occurs. The road link, which acts as a service facility to the vehicles, can be modeled as a queue with C service channels, which is the same as the capacity C. In effect, our queueing model is an M/G/C/C queue with state-dependent service rate.

Because the arrival rates of vehicles are independent of the number of vehicles already present on the road link, in our case, we can define the arrival

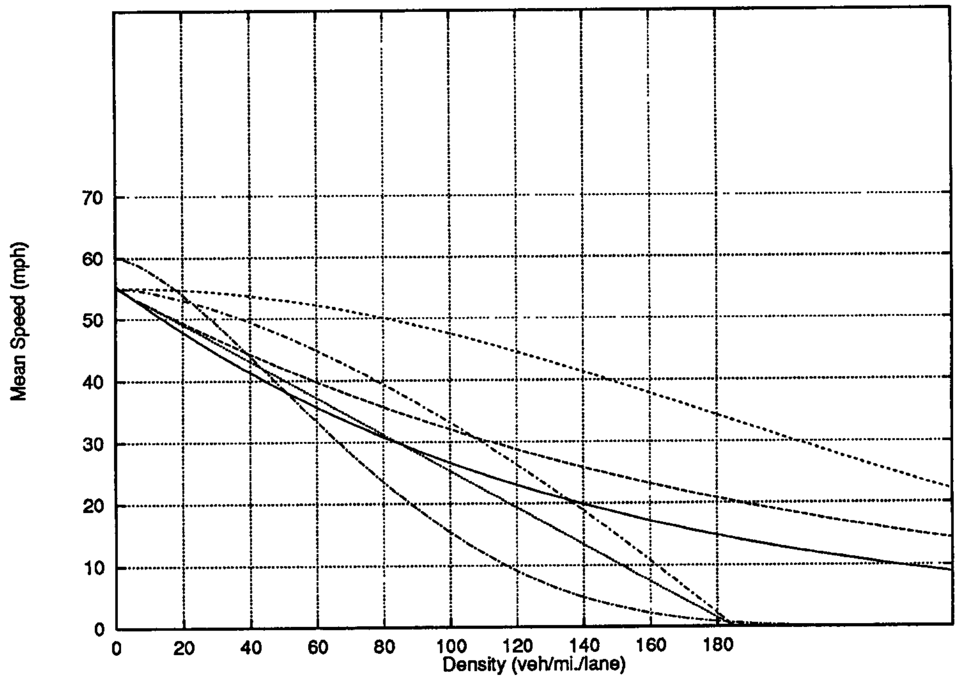

Fig. 2. Exponential approximation of empirical traffic flow models at mean speed = 55 mph. MGCC-EXP (——); Underwood (– – –); Edie (- - - - -); Greenshields (· · · · · ·); HCM (– · – · –); Drake (– – – – –).

rate to the system, λ, such that $\lambda = \lambda^0 = \lambda^1 = \cdots = \lambda^C$, where λ^j represents the arrival rate when there are j vehicles present on the road-link. For the M/G/C/C state-dependent queue, the steady state probabilities are expressed as follows [interested reader is referred to Cheah and Smith (1994)].

For $n = 1, 2, \ldots, C$,

$$P_n = \left[\frac{[\lambda \mathcal{S}]^n}{n!\, f(n) f(n-1) \cdots f(2) f(1)}\right] P_0$$

where

$$P_0^{-1} = 1 + \sum_{i=1}^{C} \left[\frac{[\lambda \mathcal{S}]^i}{i!\, f(i) \cdots f(2) f(1)}\right].$$

So the steady state probabilities for an M/G/C/C state-dependent queue depend on the mean service requirement, $\mathcal{S} = L/A$, and the relative service rate of each server, $f(i) = V_i/V_1$ (for $i = 1, 2, \ldots, C$).

Because the service rates of our queueing model are directly affected by congestion in the system, we use the relationships developed in the previous section for the linear congestion model and the exponential congestion model to describe $f(n)$.

The linear congestion model linearly relates the service rate of the servers in the M/G/C/C queueing model to the number of vehicles in the system. Using the equation for V_n developed earlier, we get

$$f(n) = \frac{V_n}{V_1} = \frac{A}{CV_1}(C + 1 - n) = \left(\frac{C + 1 - n}{C}\right).$$

Using this expression for $f(n)$ in the state probabilities equations, we obtain

$$P_n = \left[\frac{[\lambda L/A]^n}{\prod_{j=1}^{n} j[(C + 1 - j)/C]}\right] P_0$$

where

$$P_0^{-1} = 1 + \sum_{i=1}^{C} \left[\frac{[\lambda L/A]^i}{\prod_{j=1}^{i} j[(C + 1 - j)/C]}\right].$$

Note that L is expressed in miles and λ in hr^{-1}.

The exponential congestion model provides an exponential relationship between the service rate of servers in the M/G/C/C queue and the number of vehicles in the system. Again, using the travel speed equation developed earlier for the exponential congested model, we get

$$f(n) = \frac{V_n}{V_1} = \frac{A}{V_1}\exp\left[-\left(\frac{n-1}{\beta}\right)^{\gamma}\right] = \exp\left[-\left(\frac{n-1}{\beta}\right)^{\gamma}\right].$$

Substituting this expression for $f(n)$ in the state probabilities, we obtain

$$P_n = \left[\left(\frac{\lambda L}{A} \right)^n \bigg/ \prod_{j=1}^{n} j \exp\left[-\left(\frac{j-1}{\beta} \right)^{\gamma} \right] \right] P_0$$

where

$$P_0^{-1} = 1 + \sum_{i=1}^{C} \left[\left(\frac{\lambda L}{A} \right)^i \bigg/ \prod_{j=1}^{i} j \exp\left[-\left(\frac{j-1}{\beta} \right)^{\gamma} \right] \right].$$

It should be noted here that these steady-state probabilities can be used to compute performance measures such as the *probability of balking* (P_{balk} is obtained by using P_n when $n = C$ where C is the capacity of the node) and *throughput* (Θ). Throughput from a node i is computed as $\Theta = \lambda_{\text{eff}}(1 - P_{\text{balk}})$, where λ_{eff} is the *effective arrival rate* to node i. Also, the state-dependent M/G/C/C queue is *stochastically equivalent* to the state dependent M/M/C/C queue provided that the mean service rates of the two queues are equal [the interested reader is referred to Cheah and Smith (1994)].

3.1 Single Link With Multiple Sources

The model developed above, for a single source generating flow, can be easily extended to the case when customers arrive from multiple sources to a single road link of length L, N lanes, and a jam density, k. Let $\lambda_1, \lambda_2, \ldots, \lambda_J$ be the arrival rates of customers from J different sources who have to travel $L_1, L_2, \ldots, L_J$ miles, respectively, on the road link. This case can be modeled as a new single road link of length L' having the same number of lanes (N) and jam density (k) with the new arrival rate λ' where

$$L' = \frac{\Sigma_{i=1}^{J} \lambda_i L_i}{\Sigma_{i=1}^{J} \lambda_i} \quad \text{and} \quad \lambda' = \sum_{i=1}^{J} \lambda_i.$$

Thus, λ' is the sum of all the individual arrival rates from J different streams, and L' is the average distance traveled by all the arrivals from all of the different sources, calculated as a weighted average according to the different individual arrival rates $\lambda_1, \lambda_2, \ldots, \lambda_J$.

This new single link can now be solved using the linear and/or exponential analytical congestion model presented above for a single road link by replacing the original λ and L with λ' and L' respectively.

3.2 Simulation Experiments

In this section, in order to verify the analytical results obtained so far, we compare the performance measures obtained from the linear and exponential analytical congestion models with simulation models. The simulation programs simply represent an M/G/C/C state-dependent queue which was used to obtain the queueing performance measures. The linear and exponential congestion models were coded into a FORTRAN program called TRLEAM, which calculates the values of the balking probability when arriving vehicles find the system saturated P_{balk} (where $P_{\text{balk}} = P_C$), average time in system, average customers waiting in system, and the throughput rate Θ, given λ, L, and N. The simulation models were developed using SIMAN IV. Both were run on VAX machines operating under the VMS Ver 6.1 operating system.

The performance measures used for comparison purposes are measures commonly used in queueing theory (the definitions and derivations of such measures can be found in any queueing theory book (Gross and Harris (1985))) and are stated as:

1. Average time spent by the vehicles in the given road segment.
2. Average number of vehicles utilizing the segment.
3. The probability of balking when entities arrive to find the segment at capacity (P_{balk}).
4. The throughput rate (Θ) which is defined as the number of vehicles departing the roadlink per time unit. It is noted here that Θ can be interpreted as: $\Theta = \lambda_{\text{eff}} = \lambda(1 - P_{\text{balk}})$.

The experiments carried out were for unidirectional flow on a single road segment. The experiments were run under different values of physical characteristics (length, number of lanes, capacity) of the road segment, and for varying values of arrival rate (λ), jam density (k) and free flow speed ($A = V_1$). Tables I and II summarize the results obtained from analytical models and simulations performed for fixed free flow speeds of 55 and 60 mph, respectively. It can be seen that the performance measures obtained from analytical congestion models were very close to the values obtained using simulation. Furthermore, careful observation reveals that the exponential congestion model provides a better approximation for the performance measures than does the linear model. The linear congestion model performs satisfactorily under light traffic and higher capacity conditions but deteriorates under heavy traffic or low capacity conditions (as can be observed in the anomalies occurring in Tables I and II). This is consistent with the linear congestion curve fitted to the empirical curves. The linear curve drops rapidly to the jam density, whereas the exponential curve decays slowly.

TABLE I

Comparison of Simulation and Analytical Results for Speed = 55 mph

λ (vph)	L (mi)	N (lanes)	k (veh/mi-lane)	Model	TIS (hrs)	NIS (veh.)	P_{balk}	Throughput (Θ)
1000	1	1	220	Simulation	0.02112	20.78002	0	1000
				MGCC-EXP	0.021	21.178	0	1000
				MGCC-LIN	0.020	20.012	0	1000
2000	1	1	185	Simulation	0.02683	53.49253	0	2000
				MGCC-EXP	0.027	53.742	0	2000
				MGCC-LIN	3.266	184.970	0.97168	56.64
	1	1	220	Simulation	0.02685	53.60046	0	2000
				MGCC-EXP	0.027	53.742	0	2000
				MGCC-LIN	0.026	50.618	0.025239	1949.522
	1	1	265	Simulation	0.02683	53.49253	0	2000
				MGCC-EXP	0.027	53.742	0	2000
				MGCC-LIN	0.022	43.567	0	2000
	1	2	185	Simulation	0.02107	42.00707	0	2000
				MGCC-EXP	0.021	42.152	0	2000
				MGCC-LIN	0.020	40.901	0	2000
	1	2	265	Simulation	0.02107	42.00707	0	2000
				MGCC-EXP	0.021	42.152	0	2000
				MGCC-LIN	0.020	39.281	0	2000
	1	3	220	Simulation	0.01991	39.70886	0	2000
				MGCC-EXP	0.020	39.837	0	2000
				MGCC-LIN	0.019	38.628	0	2000
	1	3	265	Simulation	0.01991	39.70886	0	2000
				MGCC-EXP	0.020	39.837	0	2000
				MGCC-LIN	0.019	38.202	0	2000
4000	0.25	1	200	Simulation	0.01786	47.87804	0.3279229	2688.3084
				MGCC-EXP	0.018	47.876	0.329822	2680.712
				MGCC-LIN	0.213	49.933	0.9415	233.998
	1	1	220	Simulation	0.08558	209.68657	0.36105577	2555.77692
				MGCC-EXP	0.089	218.392	0.386	2455.077
				MGCC-LIN	3.944	219.986	0.9861	55.782

TIS, average time in system; NIS, average number in system.

It must be noted here that the simulation programs even for a single node (modeled as a state-dependent queue) and relatively small numerical values took an excessive amount of CPU time and disk space. A single road link model of dimensions $L = 1$ mile, $N = 1$, with $k = 185$ veh/mi-lane and $\lambda = 2000$/hr, when simulated for 20 time units, took approximately 44 hours of CPU time with an I/O count of over (1.7×10^7). On the other hand, analytical FORTRAN programs for larger systems with higher traffic volumes ran in under a second.

3.3 Model Validation

In this section, we attempt to validate our congestion models by undertaking experiments based on field data documented in a report published by Federal Highway Administration (1976).

This set of field data was collected along Santa Monica Freeway (East- and Westbound) at two different time points (Federal Highway Administration, 1976) by dividing the length of the freeway into "*Stations*." Using the input parameters for this data set in our program TRLEAM, we were able to obtain and compare the performance measure from field data (Average Number of Vehicles in the System) with our linear and exponential congestion models. The results are summarized in Table III. It was found that the linear and exponential congestion models provide reasonably close values for the performance measure for some of the stations. In all of the other cases, though, it can also be seen that the congestion models underestimate the values of the performance measure.

It should be noted here, that finding studies containing "complete" field data information proved to be an extremely difficult task. While some studies provided us with input parameters (length, lanes, jam density, speed, arrival rate), they lacked information regarding the output/performance measures (average number in system, average time spent in

TABLE II

Comparison of Simulation and Analytical Results for Speed = 60 mph

λ (vph)	L (mi)	N (lanes)	k (veh/mi-lane)	Model	TIS (hrs)	NIS (veh.)	P_{balk}	Throughput (Θ)
1000	1	2	220	Simulation	0.01885	18.78383	0	1000
				MGCC-EXP	0.019	18.859	0	1000
				MGCC-LIN	0.017	17.353	0	1000
2000	1	1	185	Simulation	0.02859	57.01717	0	2000
				MGCC-EXP	0.029	57.306	0	2000
				MGCC-LIN	2.963	184.934	0.968795	62.408
	1	2	185	Simulation	0.02114	42.15150	0	2000
				MGCC-EXP	0.021	42.305	0	2000
				MGCC-LIN	0.019	37.057	0	2000
	1	2	220	Simulation	0.02114	42.15150	0	2000
				MGCC-EXP	0.021	42.305	0	2000
				MGCC-LIN	0.018	36.343	0	2000
	1	3	220	Simulation	0.01954	38.96945	0	2000
				MGCC-EXP	0.020	39.101	0	2000
				MGCC-LIN	0.018	35.216	0	2000
4000	1	2	220	Simulation	0.02828	112.85772	0	4000
				MGCC-EXP	0.028	113.275	0	4000
				MGCC-LIN	0.021	82.007	0	4000

TIS, average time in system; NIS, average number in system.

system, flow while departing the segment, probability of balking/blocking, etc.), and vice-versa. The field data study used in this paper, while providing most of the required input/output values, provided the values for *observed densities* (or perhaps, optimal densities) rather than jam densities. Hence, we were forced to use observed densities in our computations while comparing the performance of our congestion models with the field data. Another key point worth mentioning while validating theoretical models with field data, is that several varying "factors" (such as road conditions, weather, visibility, etc.) can affect the traffic flow and collection of data at any given segment of freeway. So no singular theoretical generalized model could be expected to accurately compare with the field data collected at any given time.

3.4 Example Problem

In this section, we demonstrate, by way of a simple example, how the analytical congestion models developed earlier can be used to obtain performance measures for vehicular traffic networks. Consider the simple 11-node network shown in Figure 3, obtained by transforming some original road network by using the method described earlier.

Each node in this network represents a single road-link having some dimensional features and can be modeled as an M/G/C/C state-dependent queue where the capacity, C (= number of service channels), is calculated using capacity formula derived earlier in Section 3. Vehicles arrive to this network as a Poisson process with some rate λ at node 1 and depart the system from node 11 with a throughput rate Θ.

Several design and control issues arise while looking at this network. Some of the issues are:

- What is the shortest route from node 1 to node 11? What are the k-shortest paths?
- On an average, how much time is spent by a vehicle while traversing this network?
- How many vehicles can depart per time unit, and what is the utilization of each road-link?
- Could changes in dimensional features of some of the road-links (such as increasing/decreasing the number of lanes, controlling the number of vehicles allowed to enter the system, preestablished routing for certain percentage/type of vehicles, etc.) lead to improvement in the performance of this network?

Because the analytical congestion models developed in this paper are for a single road link, we can solve this example network by studying each node independently. It must be mentioned here that we are aware of possible blocking occurring between the links in the network which cannot be appropriately included (for optimization purposes) while computing the performance of the whole network unless further model refinements are imposed (See Cheah and Smith (1994) for one way to accommodate the blocking between the links using our M/G/C/C

TABLE III

Santa Monica Freeway East and Westbound at Time = 07:01:00 and 07:01:20

Station Location	Length *L* (mi)	Lanes *N*	Arrival Rate λ (vph)	Jam Density (veh/mi-lane)	Speed (mph)	Model	NIS (veh)
SM16E	0.47	4	7200	33.91	53.17	Field Data	60
						MGCC-LIN	62.98
						MGCC-EXP	59.82
	0.47	4	7080	33.29	52.46	Field Data	59
						MGCC-LIN	61.98
						MGCC-EXP	58.74
SM17E							
Lane 1	0.60	1	1980	37.96	52.46	Field Data	33
						MGCC-LIN	21.95
						MGCC-EXP	19.78
Lane 3	0.60	1	1680	32.98	51.74	Field Data	28
						MGCC-LIN	18.94
						MGCC-EXP	16.67
SM18E							
Lane 2	0.44	1	1980	38.58	51.58	Field Data	33
						MGCC-LIN	15.93
						MGCC-EXP	14.02
SM17W							
Lane 1	0.60	1	1320	25.82	55.58	Field Data	22
						MGCC-LIN	14.91
						MGCC-EXP	12.75
SM17E							
Lane 1	0.60	1	2040	40.13	51.19	Field Data	34
						MGCC-LIN	23.95
						MGCC-EXP	21.42
Lane 3	0.60	1	1560	31.11	50.66	Field Data	26
						MGCC-LIN	17.94
						MGCC-EXP	15.56
SM18E							
Lane 2	0.44	1	2040	39.51	51.91	Field Data	34
						MGCC-LIN	16.93
						MGCC-EXP	14.91
SM17W							
Lane 1	0.60	1	1380	27.38	50.22	Field Data	23
						MGCC-LIN	15.93
						MGCC-EXP	13.61

NIS, average number in system.

model). However, we feel that the purpose of this example is to show the usage of our analytical models as a fundamentally new approximation tool for studying vehicular traffic flows.

With the assumption of node independence, several routes can be generated deterministically, and the performance measures for the routes be computed for each node using the analytical congestion models developed in this paper. The vehicles, after entering the network at node 1, traverse the nodes on the designated route, and depart from the network through node 11. The capacities of the nodes can be varied (for example, by varying the number of lanes or the arrival rate to node 1) to study the effects on performance measures (see Section 4 for a detailed discussion). For each route, performance measures such as the average time spent in the system by the vehicles, the throughput rate, and utilization factor for each node can be calculated. All calculations are done using the FORTRAN programs written for our analytical congestion models (refer to Section 3.2). Such analyses, using analytical congestion models, can provide valuable insight

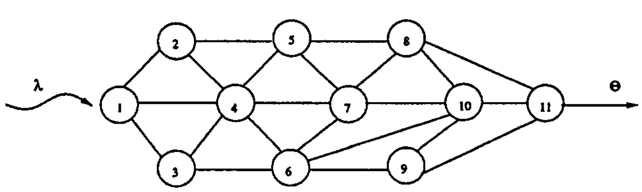

Fig. 3. Example network.

about the performance of the road-network and can be helpful in design and control of the process.

4. OPTIMIZATION AND SENSITIVITY ANALYSIS

IT IS INTERESTING to study how the variation in the values of certain variables affects the values of some other variables. In the context of roadway systems, one could be interested in finding that value of the arrival rate λ which maximizes the throughput (Θ) for a given link. Another interesting issue is that, given the arrival rate λ and fixing one, and only one, of the other dimensions (either L or N), what effect do the variations in the values of the unfixed variable have on the blocking/balking probability (P_{balk}). This kind of analysis helps in determining the sensitivity of design and control variables of the system with respect to the behavior of the remaining variables. In the next subsections, we attempt to address these issues.

4.1 λ Analysis

The two issues explored in this section are the effects of arrival rate (λ) on the balking probability and the system throughput. Given dimensions of the roadway segment and an upper bound on the blocking probability, is there a value of λ that maximizes the throughput? Which value of λ can be used to achieve the upper bound on P_{balk}?

The question of finding that λ which maximizes throughput is of practical importance because knowledge of λ^* at which maximum throughput is achieved helps in making routing decisions, because we can accordingly control the arrival stream to the road link, thereby achieving overall optimization. Using the program TRLEAM to compute the values of P_{balk}, we can calculate the throughput (Θ) as

$$\Theta = \lambda(1 - P_{\text{balk}})$$

As the arrival rate to the network increases, the probability of balking customers also increases, with resulting changes in Θ. By plotting curves such as those shown in Figures 4–7, we are able to find the value of λ^* at which Θ is maximized. It can be seen by comparing the linear congestion curves (Figs. 4 and 5) and exponential congestion curves (Figs. 6 and 7) that the value of λ drops more rapidly in linear congestion than in exponential case once λ^* is achieved.

As mentioned earlier, TRLEAM can be used to compute the value of the blocking probabilities for given dimensions of L and N. By plotting the curves of constant capacity (see Figs. 8 and 9) for the road link with $\rho = \lambda/\mu_1$ on the x-axis and P_{balk} on the y-axis, we can determine the value of λ. Since the value of P_{balk} is given and μ_1 can be calculated using $\mu_1 = A/L = V_1/L$, we can compute the arrival rate as $\lambda = \rho\mu_1$.

From Figures 8 and 9 we can see that the curves with higher capacities have lower values of P_{balk}. Also, linear congestion curves attain sharper bends and at higher P_{balk} than those corresponding to the exponential congestion model. Furthermore, linear model curves approach $P_{\text{balk}} = 1$ asymptotically much faster than the exponential curves.

Another way to determine the arrival rate is by

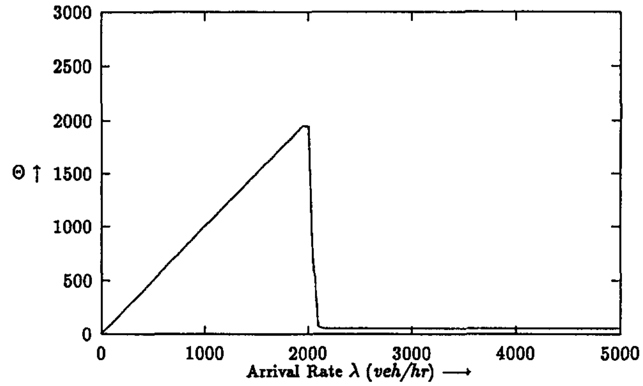

Fig. 4. Linear congestion with $L = 1$ mi and $N = 1$.

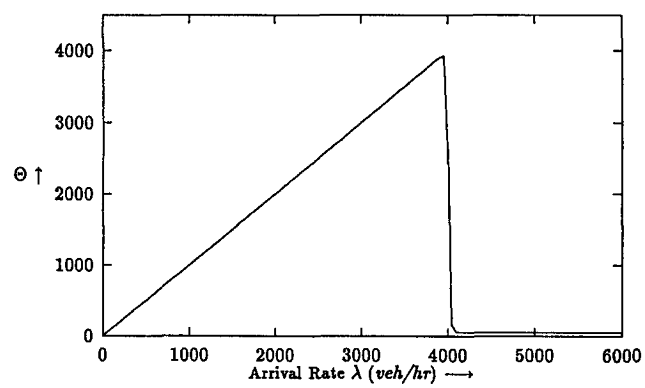

Fig. 5. Linear congestion with $L = 1$ mi and $N = 2$.

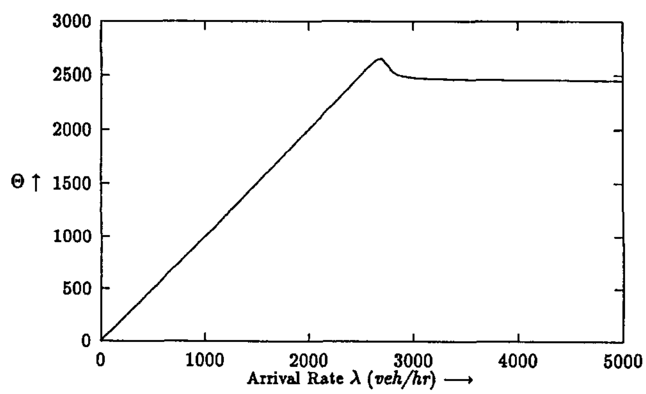

Fig. 6. Exponential congestion with $L = 1$ mi and $N = 1$.

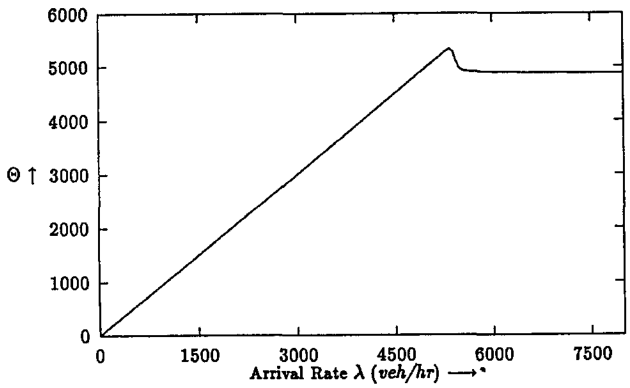

Fig. 7. Exponential congestion with $L = 1$ mi and $N = 2$.

using a FORTRAN program called TRLESLAM. Given the values of the road link length L, the number of lanes N, and the desired upper bound on the balking probability, TRLESLAM calculates the optimal value of the arrival rate, λ, by using bisection as a unidimensional search technique over many values of λ.

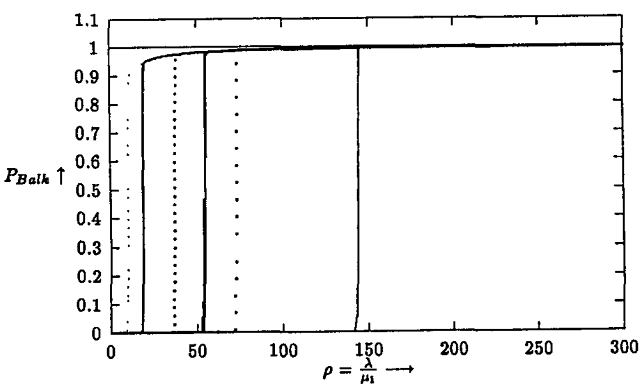

Fig. 8. Linear congestion with varying capacities - ρ vs. P_{balk}. $P_{\text{balk}} = 1$ (——); $C = 55$ (⋯⋯); $C = 110$ (——); $C = 220$ (• • • • •); $C = 330$ (——); $C = 440$ (• • •); $C = 880$ (——).

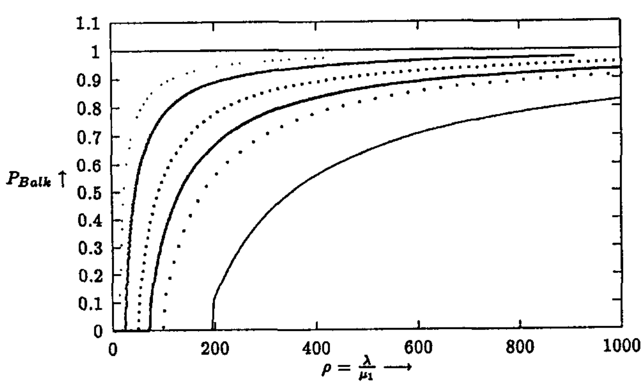

Fig. 9. Exponential congestion with varying capacities - ρ vs. P_{balk}. $P_{\text{balk}} = 1$ (——); $C = 55$ (⋯⋯); $C = 110$ (——); $C = 220$ (• • • • •); $C = 330$ (——); $C = 440$ (• • •); $C = 880$ (——).

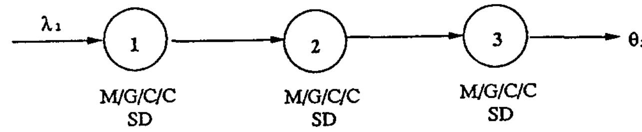

Fig. 10. Three nodes in series topology.

4.2 $\mathcal{N}$-Variations

Another variable of interest is the number of lanes on the road segment (N). Provided the arrival rate for a segment has been predetermined, and the values of L and P_{balk} are fixed, we can study the effects of increasing or decreasing the number of lanes on system performance. This variation affects the road-link capacity and hence affects the congestion experienced on the road-link, which is useful to make design and construction decisions of whether or not one or more lanes need to be added to help ease congestion.

A FORTRAN program named TRLESN, available from the authors, can calculate the optimal value for number of lanes given the arrival rate, the road link length, and the desired upper bound on the balking probability.

Another method is to utilize graphs such as those shown in Figures 8 and 9. These graphs are drawn using the data output collected by running TRLEAM program. From Figures 4 and 5 for linear congestion, one can see how changes in N from one lane to two lanes lead to changes in the throughput from the link. Similar inferences can be drawn for exponential congestion from Figures 6 and 7. In both the cases, throughput appears to almost double when one lane is added.

5. BLOCKING FLOWS

SO FAR IN THIS PAPER we have developed, and studied in great detail, congestion models to capture the effects of congestion on the performance of a single road segment. This allowed us to explore each road segment independently and to compute performance measures of each node independently as well. When more than one road segments are studied together, the assumption of independence is no longer valid. A saturated downstream node affects the performance of the upstream node. In this situation, the downstream node is called the *bottleneck node*. For illustrative purposes, Figure 10 presents three road segments in a series topology. In this scenario, if the downstream node 3 experiences blocking, either due to heavy traffic or low capacity, it will have a negative impact on the performance of the upstream nodes 1 and 2. How can this blocking be accommodated in the context of our modeling approach? In this section, we briefly present some of our current work useful in

TABLE IV

Simulation versus Analytical for Three Node Series Topology

λ vph	L^i mi	N^i veh/mi-lane	K^i mph	Speed^i	Model	NIS^1	P^1_{balk}	NIS^2	P^2_{balk}	NIS^3	P^3_{balk}
1999	(0.5, 0.5, 0.5)	(4, 4, 4)	(185, 185, 185)	(65, 65, 65)	Simu	16.35	0.50	8.14	0.48	7.20	0.00
					Exp	16.26	0.49	8.07	0.49	7.16	0.00
1999	(0.75, 0.75, 0.75)	(3, 3, 3)	(210, 210, 210)	(58.5, 58.5, 58.5)	Simu	28.10	0.00	28.10	0.00	28.10	0.00
					Exp	27.92	0.00	27.92	0.00	27.92	0.00
2999	(0.5, 0.5, 0.5)	(4, 4, 4)	(185, 185, 185)	(65, 65, 65)	Simu	25.24	0.49	12.76	0.48	10.81	0.49
					Exp	250.17	0.47	12.70	0.48	10.79	0.49
2999	(0.75, 0.75, 0.75)	(3, 3, 3)	(210, 210, 210)	(58.5, 58.5, 58.5)	Simu	44.12	0.00	44.12	0.00	44.12	0.00
					Exp	44.08	0.00	44.08	0.00	44.08	0.00
5999	(0.5, 0.5, 0.5)	(4, 4, 4)	(185, 185, 185)	(65, 65, 65)	Simu	56.40	0.50	270.26	0.48	17.21	0.48
					Exp	56.27	0.47	270.12	0.48	17.11	0.47

NIS, average number in system.

addressing the issue of the effect of downstream bottleneck nodes on the upstream nodes.

For simple queueing networks, exact analytical methods can usually be applied to obtain the necessary performance measures. However, if a queueing network model does not possess a product form solution, then one relies on approximation approaches. The complexity of M/G/C/C state dependent queueing models makes it impossible to use direct analytical methods. The approximation approach undertaken in our research, for the purposes of analyzing more than one road segment (queueing node), is called the *generalized expansion method*, which was first developed by KERBACHE and SMITH (1987). The *generalized expansion method* is a robust and effective approximation approach which is characterized to be a combination of Repeated Trials and Node-by-Node decomposition solution procedures. This method expands the original queueing network by introducing an artificial holding node preceding each blocking node. A detailed discussion on the intricacies of the *generalized expansion method* can be found in Kerbache and Smith (1987). The current work uses a set of simultaneous nonlinear equations generated by the *generalized expansion method* to capture the effect of blocking and to obtain the performance measures of all the road segments for the vehicular traffic flow problem.

In Table IV, we present a sample of the results obtained for a 3-node series topology where each road segment has identical dimensions with the arrival rates to the first node taking on the values of $\lambda = 1999$ vph, 2999 vph, and 5999 vph. The performance measures for each road segment were computed using FORTRAN program called SERIES which solves the set of simultaneous nonlinear equations generated by the *generalized expansion method* for tandem queues. The analytical numerical results were compared with simulation runs performed using SIMAN IV. Both simulation and analytical experiments were run on VAX machines operating under VMS Ver 6.1 operating system. The performance measures used for comparison were:

1. Average number of vehicles utilizing road segment.
2. The probability of balking at each road segment when entities arrive to find segment i at capacity (P^i_{balk}).

As can be seen from the tabulated experiments, the analytical model using *generalized expansion method* provides very close results to the simulation output. It appears that, in almost all of the cases, the analytical approach slightly underestimates the performance measures. Furthermore, as we had observed in the simulation experiments for a single road link, the simulation experiments for the 3-node series topology took an excessive amount of CPU time (an average of five days worth of CPU time) and disk space while the analytical programs using FORTRAN ran in under a second. The same technique is being employed by the authors to analyze split and merge topologies as well, and more detailed results of this work will appear in future publications.

6. SUMMARY AND CONCLUSIONS

IN THIS PAPER, we studied the dynamic and stochastic effects of congestion on vehicular traffic flow by using M/G/C/C state-dependent queueing models. Linear and exponential models with state dependent service rate, depending on the number of customers in the system, were developed to account for the decay in the service rate of a unidirectional single road segment. These analytical models provide a fundamentally new tool to an analyst for not only calculating the performance measures for road segments, but also for making design and control decisions for roadways. Finally, computer programs written in FORTRAN-77, for implementation of the mathematical models described in the paper, are available upon request from the authors.

Further extensions of this work include modeling traffic intersections as merge and split topologies, studying networks under heavy congestion where blocking between adjacent traffic segments occurs, and routing within vehicular evacuation networks.

REFERENCES

A. S. ALFA AND M. F. NEUTS, "Modeling Vehicular Traffic using the Discrete Time Markovian Arrival Process," *Transp. Sci.* **29:2,** 109–117 (1995).

M. C. BELL, *Measures of Queueing Performance for a Traffic Network*, University of Newcastle upon Tyne, Newcastle, England, 1980.

D. S. BERRY, "Other Methods for Computing Capacity of Signalized Intersections," *Transp. Res. Board Presentation* (1977).

J. CHEAH, State Dependent Queueing Models, Master's thesis, Department of Industrial Engineering and Operations Research, University of Massachusetts, Amherst, MA 01003, 1990.

J. CHEAH AND J. M. SMITH, "Generalized M/G/C/C State Dependent Queueing Models and Pedestrian Traffic Flow," *Queueing Syst.* **15,** 365–386 (1994).

R. W. CONWAY AND W. L. MAXWELL, "A Queueing Model with State Dependent Service Rates," *J. Indust. Eng.* **12,** 132–136 (1962).

J. S. DRAKE, J. L. SCHOFER, AND A. D. MAY, JR., "A Statistical Analysis of Speed Density Hypotheses," in *3rd International Symposium on the Theory of Traffic Flow Proceedings*, Elsevier North Holland, Inc., New York (1967).

S. M. EASA AND A. D. MAY, Generalized Procedure for Estimating Single- and Two-Regime Traffic-Flow Models, Transportation Research Board, Record 772, TRB, Washington, D.C., 24–37 (1980).

L. C. EDIE, "Following and Steady-State Theory for Non-Congested Traffic," *Oper. Res.* **9,** 66–76 (1961).

M. ELDOR AND A. D. MAY, Selected References on Application on Computer Simulation to Transportation Systems, *ITTE Library Reference 43*, University of California, Berkeley, California (1975).

FEDERAL HIGHWAY ADMINISTRATION, Freeway Data For Incident And Nonincident Conditions, Report No. FHWA-RD-76-175, 1, Offices of Research and Development, Washington, D.C. (1976).

P. FOX AND F. LEHMAN, "Digital Computer Simulation of Automobile Traffic," *Traffic Q.* 21:1, 53–66 (1967).

R. E. FRANKLIN, "The Structure of a Traffic Shock Wave," *Civil Eng. Publ. Works Rev.* **56:662,** 1186–1188 (1961).

B. D. GREENSHIELDS, "A Study in Highway Capacity," *Proc. Highway Res. Board,* **14,** 458 (1935).

D. GROSS AND C. M. HARRIS, *Fundamentals of Queueing Theory*, John Wiley and Sons, New York 1985.

F. S. HILLIER, R. W. CONWAY, AND W. L. MAXWELL, "A Multiple Server Queueing Model with State Dependent Service Rates," *J. Indust. Eng.* **15,** 153–157 (1964).

L. KERBACHE AND J. M. SMITH, "The Generalized Expansion Method for Open Finite Queueing Networks," *Eur. J. Oper. Res.* **32,** 448–461 (1987).

M. KUWAHARA AND G. F. NEWELL, Queue Evolution on Freeways Leading to a Single Core City during the Morning Peak, *10th Int. Symp. Transp. Traffic Theory Proc.* 21–40 (1987).

A. D. MAY, *Traffic Flow Fundamentals*, Prentice Hall, Englewood Cliffs, NJ 1990.

A. D. MAY AND H. E. M. KELLER, "A Deterministic Queueing Model," *Transp. Res.* **1:2,** 117–128 (1967).

K. MOSKOWITZ AND L. NEWMAN, Traffic Bulletin No. 4—Notes on freeway capacity, Highway Research Board, Record 27, HRB, Washington D.C., 44–68 (1963).

G. F. NEWELL, "Approximation Methods for Queues with Application to the Fixed-Cycle Traffic Light," *SIAM Rev.* **7:2,** 223–240 (1965).

L. A. PIPES, "Wave Theories of Traffic Flow," *J. Franklin Inst.* **280:1,** 23–41 (1965).

P. I. RICHARDS, "Shock Waves on the Highway," *Oper. Res.* **4:1,** 42–51 (1956).

J. M. SMITH, "State Dependent Queueing Models in Emergency Evacuation Networks," *Transp. Res. B* **25B:6,** 373–389 (1991).

TRANSPORTATION RESEARCH BOARD, The Application of Traffic Simulation Models. Transportation Research Board, Special Report 194, TRB, Washington, D.C., 1981.

TRANSPORTATION RESEARCH BOARD, Highway Capacity Manual, Special Report 204, Washington D.C., 1985a.

TRANSPORTATION RESEARCH BOARD, Highway Capacity Manual, Special Report 209, Washington D.C., 1985b.

R. T. UNDERWOOD, "Speed, Volume, and Density Relationships, Quality and Theory of Traffic Flow," *Yale Bur. Highway Traffic*, New Haven, Connecticut, 141–188 (1961).

M. R. WIGAN, "Applications of Simulation to Traffic Problems—An Annotated Bibliography," *Road Res. Bibliog. 102/MRW*, Transport and Road Research Laboratory, Crowthorne, England (1969).

S. C. WIRASINGHE, "Determination of Traffic Delays from Shock Wave Analysis," *Transp. Res.* **12,** 343–348 (1978).

S. YUHASKI AND J. M. SMITH, "Modeling Circulation Systems in Buildings using State Dependent Queueing Models," *Queueing Syst.* **4,** 319–338 (1989).

(Received: January 1996; revisions received: May 1996, August 1996; accepted: September 1996)