Distribution-Free Model for Estimating Random Queues in Signalized Networks

ANDRZEJ TARKO AND NAGUI ROUPHAIL

A general-arrival, bulk service time queueing model is formulated for studying the distribution of random queues in signalized networks. The model is predicated on the occurrence of three traffic stream transformations: merging, splitting, and filtering. The model is applied to steady-state conditions (traffic intensity < 1.0) but can be eventually converted to a time-dependent form to account for oversaturation effects. A comparison of the results of the model with those of comparable models in the literature confirms that the use of random queue estimates derived from the assumption of a Poisson arrival process is inappropriate for networks. Marginal adjustments to the Poisson process by including a variance-to-mean ratio of the departure distribution improve the random queue estimate to a point. The results also confirm recent observations by Newell about the relationship of stochastic queues in an arterial network with their counterparts at isolated intersections. In general queue estimates for the network case are substantially smaller than those incurred at an isolated intersection with similar traffic intensity. The difference is attributable primarily to the process of traffic filtering.

Vehicle delays at signalized intersections contribute substantially to travel times on an urban street network. Delay is now the basic criterion for evaluating the level of service (LOS) at signalized intersections and a key ingredient for evaluating the LOS on arterials (*1*). The ability of a traffic analyst to estimate vehicle delay is critical in evaluating advanced traffic management systems (ATMS) as well as quantifying the environmental consequences of traffic decisions.

Average vehicle delay at a signalized intersection can be expressed as the sum of nonrandom and overflow delay components. Nonrandom delay refers to the average vehicle delay experienced with the assumption that traffic demand is uniform and averaged over all cycles during the analysis period. Overflow delay encompasses the additional delay caused by the randomness in arrival headways within each cycle and from one cycle to the next, in addition to that incurred when flow exceeds capacity for some period of time. Within-cycle random variations are usually negligible in terms of their impact on delay, an effect that is also not considered in this paper. Thus, the residual queue remaining at the end of the green phase (herein denoted as N_o) is considered the only source of overflow delay. The relationship between average overflow queue and average random delay d_o can be approximated for steady-state conditions as follows (*2*):

$$d_o = \frac{N_o}{q} \tag{1}$$

where

d_o = random delay (sec),
N_o = random queue (veh), and
q = arrival rate (veh/sec).

Since this relationship is straightforward and independent of the queueing model distributions, random delay is often investigated through the estimation of random queues.

BACKGROUND

Nonrandom delay formulas exist for both isolated and coordinated intersections (*1,3*). Estimating the second delay component for a signalized network is still a challenging research issue. Earlier theoretical work on queueing theory (*4–6*) hints at some major difficulties in obtaining delay formulas for general arrival and departure distributions. The most general steady-state delay models have been derived by Darroch (*5*), Newell (*2*), and McNeil (*7*), who incorporate the variance-to-mean ratio I_a in their models to include binomial or compound Poisson arrivals (Darroch's processes). Since these works did not deal directly with signalized networks, these questions remained: what actually are the arrival processes in signalized networks? and how is the value of I_a estimated if Darroch's processes are appropriate for signalized networks?

Van As addressed these issues using the Markov chain approach to model delays and arrivals at two closely spaced signals (*8*). He concludes that the Miller model improves random delay estimation for signalized networks in comparison with the Webster model. However, Van As's results also indicated that Miller's formula overestimated random delay in some cases. It is unclear whether that bias was caused by the non-Darroch's arrival process or by the coordinate transformation technique (*9*) used to obtain the time-dependent models investigated by Van As.

Tarko et al. have investigated the impact of an upstream signal on random delay using cycle-by-cycle macrosimulation (*10*). They found that in some cases the ratio I_a does not properly represent the non-Poisson arrival process and generally overestimates delay. The additional weakness of such models lies in the estimation of I_a. Although Van As worked out a straightforward formula for I_a, its dependence on the I_a calculated at an upstream signal creates the possibility of a systematic error propagation problem in the course of the calculations. To avoid that problem, Tarko et al. (*10*) proposed a random delay model that uses a function of the capacity differential between the critical upstream and subject signals instead of the I_a ratio to improve delay estimation in signalized networks. Their work also confirms that traffic platooning—that is, signal progression—in a signalized network operating on a common signal cycle has no effect on the cycle-to-cycle variation of the arrival

A. Tarko, Urban Transportation Center, University of Illinois at Chicago, 1033 West Van Buren, Suite 700 South, Chicago, Ill. 60607. N. Rouphail, Civil Engineering Department, North Carolina State University, P.O. Box 7908, Raleigh, N.C. 27695.

distribution process. In other words, signal offset does not affect random delays or queues.

The present paper can be seen as an extension of the author's previous work (*10*). A bulk service queueing model is presented that enables the description of the distribution of vehicle arrivals, departures, and random queues in a signalized network. The model is evaluated by comparing it with well-recognized random queue models for isolated intersections [Khintczine-Pollaczek, Newell (*2*), and Akçelik (*11*)] and for networks [modified Newell (*12*), Miller (*13*), and Tarko-Rouphail (*10*)]. Furthermore, Newell's hypothesis on the average random queue along a signalized arterial (*12*) is tested. Finally, a sensitivity analysis on the effect of secondary flows (midblock and turning movements) on random queues is presented and discussed.

ARRIVAL DISTRIBUTION IN A SIGNALIZED NETWORK

Consider an urban street network on which most intersections are signalized. An additional, and reasonably valid, assumption is that all these signals operate on a common signal cycle. The traffic stream moving through the network is subject to the following transformations: it can

- Merge with other traffic streams,
- Split into separate traffic streams, or
- Be filtered by traffic signals.

In such transformations, a traffic stream is represented by its arrival distribution in time periods that are equivalent to the common signal cycle. Arrival distributions are generated at locations where a given transformation takes place. For example, consider a traffic link connecting two signalized intersections (Figure 1). The link is modeled as a sequence of cross sections at which traffic streams are merged, split, or filtered. These arrival distribution transformations are modeled using the following processes:

- *Merging* produces a combined distribution of arrivals $P(x)$ from two independent traffic streams with arrival distributions $P_1(a)$ and $P_2(a)$ as follows:

$$P(x) = \sum_{a=0}^{x} P_1(a)P_2(x-a) \tag{2}$$

In cases of three or more streams, this formula is applied consecutively, so that $P_1(a)$ is the result from the previous application and $P_2(a)$ corresponds to the next stream to be combined.

- *Splitting* produces a distribution $P_s(x)$ of arrivals drawn with probability p from a traffic stream with known arrival distribution $P(a)$ according to

$$P_s(x) = \sum_{a=x}^{A} P(a) \frac{a!}{x!(a-x)!} p^x(1-p)^{a-x} \tag{3}$$

where A is the maximum number of arrivals considered to have a finite value. For entry links into the network, the Poisson distribution may usually be applied to estimate the number of arrivals. In this case the value of A is set sufficiently large to neglect the truncation error.

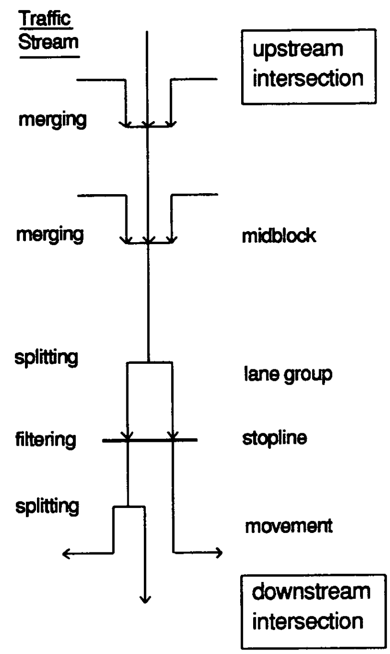

FIGURE 1 Traffic streams on a link in random queue modeling.

- Signal *filtering* transforms the arrival distribution $P(x)$ just upstream of an intersection into a departure distribution $P_d(x)$ just downstream of the intersection. The number of departures per cycle is equivalent to the sum of overflow queue (from previous cycles) in addition to all "new" arrivals in the subject cycle if that sum is less than the signal capacity. Otherwise the number of departures is set equal to the signal capacity

$$P_d(x) = \sum_{a=x}^{A} \Pi(a)P(x-a) \qquad \text{for } x < c \tag{4}$$

$$P_d(x) = 1 - \sum_{a=0}^{c-1} P_d(a) \qquad \text{for } x = c \tag{5}$$

where c is the fixed signal capacity per cycle and $\Pi(x)$ is the probability of x vehicles in queue at the end of green time in steady-state conditions.

Random queues on any link within the network can be handled by these transformations, first by modeling all the upstream links that feed into the subject link. For simplicity, a Poisson distribution may be assumed for arrivals on external links. Thus, the process of network link modeling is carried out in the following manner (Figure 1):

1. Combine the departure distributions from all exits at the upstream intersection.
2. Assume that the combined departure distribution in Step 1 constitutes the arrival distribution at the midblock unsignalized intersection (real or hypothetical). Combine this profile with the arrival profile from midblock traffic when applicable.
3. Assume, as in Step 2, that the combined profile at the midblock location constitutes the arrival profile at the downstream signal. Split the traffic stream in the segregation zone according to the prevailing lane assignment. For example, Figure 1 shows that through and

right-turning movements share common lanes. Thus, these movements are considered to form a single traffic stream. Left turners using exclusive lanes form a separate queue and are considered as a separate traffic stream that "splits" from the combined profile derived in Step 2.

4. Filter all separate traffic streams at the stopline. First, the random queue distribution is obtained from the arrival distribution and signal capacity. Next, the random queue distribution and the arrival distribution are used to produce the departure distribution. For example, the resulting departure profile for left turners is a final profile and may be used in modeling the appropriate downstream link.
5. Split shared traffic streams into individual movements (right turns from through traffic in Figure 1). The movements' departure distributions complete the requirements for processing the subject link.
6. Repeat Steps 1 through 5 for the downstream intersection.

This model assumes the arrival distribution to be identical to the upstream departure distribution. This assumption is valid if variations in vehicle speeds between the two intersections do not affect the number of arrivals in cycles at the downstream intersection. For long road sections, this assumption results in some underestimation of the random queues. The variations in the arrival distribution between cycles, which occur along road sections on which traffic is uninterrupted, require additional research. The arrival distribution considered here should not be confused with the average flow rate profiles used in models such as TRANSYT.

In the next section, the distribution of the random queue is modeled assuming a statistical equilibrium state. The random queue distribution is of main interest since it can be used to calculate expected random queues and delays. Splitting and merging transformations yield intermediate results that are required to either model random queues at the subject signal or continue modeling downstream links.

RANDOM QUEUE MODEL FOR GENERAL ARRIVAL DISTRIBUTIONS

A queueing system with a single server, random arrivals from a Poisson distribution, a deterministic bulk service, and queue discipline FIFO has been applied to random queue modeling at an isolated signalized intersection (*4*). The assumption of Poisson distribution is, however, too restrictive for signalized networks. Instead assume a general arrival distribution $P(x)$, where x is the number of arrivals in cycle, and overflow queue distribution in steady-state conditions $\Pi(k)$, where k is the number of vehicles in queue when green time ends. The signal capacity, expressed in number of vehicles that can be possibly served during cycle, is fixed and equal to c. An analysis of state transient probabilities under equilibrium conditions resulted in the system of balance equations for a general arrival distribution that is applicable to a signalized network:

$$\begin{aligned} &\textstyle\sum_{i=0}^{\infty} P(x \leq c - i)\Pi(i) - \Pi(k) = 0 \qquad \text{for } k = 0 \\ &\textstyle\sum_{i=0}^{\infty} P(x = c - i + k)\Pi(i) - \Pi(k) = 0 \qquad \text{for } k = 1, \ldots, \infty \end{aligned} \tag{6}$$

To avoid the trivial and infeasible solution [$\Pi(i) = 0$ for each i], the first equation in the system (Equation 6) is substituted with the constraint on steady-state queue probabilities such that all probabilities $\Pi(i)$ sum to 1:

$$\textstyle\sum_{i=1}^{n=\infty} \Pi(i) = 1 \tag{7}$$

The proposed queueing model gives an exact solution to the problem under steady-state conditions. However, to solve the equation numerically, the size of the problem must be limited to some finite and large numbers i and k, such that $i = k$. Finally, the system of linear equations in the standard form, convenient for many solution techniques, is

$$(\mathbf{a} - \mathbf{e}) \cdot \mathbf{\Pi} = \mathbf{b} \tag{8}$$

where

$\mathbf{a}$ = two-dimensional matrix ($M \times M$) with elements as follows:
$a_{1j} = 1$ for $j = 1, \ldots, M$
$a_{ij} = P(x = c + i - j)$ for $i = 1, \ldots, M$ and $j = 2, \ldots, M$
$\mathbf{e}$ = two-dimensional matrix ($M \times M$) with elements as follows:

$$e_{ij} = \begin{cases} 1 \text{ for } i > 1, j > 1, i = j \\ 0 \text{ otherwise} \end{cases}$$

$\mathbf{\Pi}$ = column vector of probabilities [$\Pi(k = 0), \Pi(k = 1), \ldots, \Pi(k = M - 1)$];
$\mathbf{b}$ = column vector with M elements [1, 0, 0 . . .]; and
M = sufficiently large number such that truncation error is negligible for solution of system (8).

In the proposed model, signal capacity is assumed to have a fixed value. Olszewski (*14*) concluded that under reasonable capacity conditions, the use of a fixed value rather than a distribution is acceptable for unopposed traffic streams. The question arises whether this finding is applicable to a signalized network, since even small variations in the upstream signal capacity are propagated downstream when filtering takes place.

To answer this question, one should recognize that the principal source of capacity variations for unopposed streams is the cycle-to-cycle variations in traffic composition. However, the traffic composition at an upstream signal tends to be replicated downstream since the same vehicles arrive at both signals with some time lag. This means that the upstream and downstream signal capacities do not vary independently, which reduces the effect of capacity variations. Estimating the random queues for two cases—fixed and independently varying capacities—yields lower and upper bounds of random queue lengths. This is beyond the scope of this work.

COMPARISON WITH EXISTING MODELS

Single Intersection Models

The first step in the model evaluation is to compare its estimates with those of several well-known models: the Khintczine-Pollaczek (K-P) second term used by Webster (*3*) and Kimber and Hollis (*9*) in their formulas, Akçelik (*11*) random delay formulas, and the Newell model (*2*) modified by Cronje (*15*). These models converted into the random queue models according to Equation 1 are presented here:

- K-P for Poisson arrivals and deterministic departure processes:

$$N_o = \frac{X^2}{2(1 - X)} \tag{9}$$

- Akçelik:

$$N_o = \frac{1.5[X - (0.67 + c/600)]}{1 - X} \quad (10)$$

- Newell (with Cronje modification) for Poisson arrivals:

$$N_o = \frac{H(\mu)X}{2(1 - X)} \quad (11)$$

where

$$H(\mu) = \exp[-(1 - X)c^{0.5} - 0.5(1 - X)^2 c] \quad (12)$$

In these equations, X is the degree of saturation, and c is cycle capacity in vehicles per cycle.

These models are compared with the authors' results and with the results obtained by Olszewski (*14*). The comparison is presented in Figure 2 for degrees of saturation 0.90 and 0.95 and for capacities varying from 10 to 120 veh/cycle. Olszewski's model based on the Markov chain produces virtually identical results to the modified Newell model and, therefore, is omitted from the comparison. The results demonstrate the significant effect of cycle capacity even for a fixed degree of saturation. The K-P model highly overestimated random queues since it does not consider the bunching of serviced vehicles during the green signal. Excellent agreement is evident between the bulk service and Newell models. Observed discrepancy between the Akçelik model and other models is a result of linearity of the first one.

Network-Based Models

The second step in model evaluation is to compare the steady-state K-P, Miller (*13*), Tarko-Rouphail (*10*), and Newell (2) with Cronje modifications (*15*) models with the bulk service model estimates. All these models can be represented using the following generalization:

$$N_o = \frac{k \cdot (X - X_0)}{(1 - X)} \quad (13)$$

where

X = degree of saturation;
k = $0.5X$ in K-P formula;
= $0.52 \times I_a \times X$ in Miller formula;
= $I_a \times H(\mu)$ in modified formula [$H(\mu)$ is calculated according to Equation 11];
= $0.408[1 - e^{-0.5(c_u - c_d)}] \times X$, where c_u and c_d are the cycle capacities for the critical upstream and subject signals, respectively, in Tarko-Rouphail formula; and
X_0 = 0 in K-P, Newell, and Miller models
= $Q_d/100$ in Tarko-Rouphail formula.

In Miller's and Newell's models, I_a is meant to incorporate the effect of non-Poisson arrivals ($I_a < 1$), and $H(\mu)$ incorporates the bulk service in Newell's model. The Tarko-Rouphail formula includes an adjustment factor as a function of signal capacities.

To provide data for comparison, a system of three signals with no turning movements is considered. Capacities for the first and second intersections were allowed to vary from 30 to 40 veh/cycle in 1-veh increments from one computation to the next. The third signal had a fixed capacity of 30 veh/cycle. Traffic volume was also fixed at an average of 27.5 veh/cycle, resulting in a fixed degree of saturation of 0.92 at the third intersection. Poisson arrivals were assumed at the first (entry) signal. Here it is recognized that the upstream signals will substantially transform the arrival pattern at the downstream intersections. Comparative results are depicted in Figure 3 for the second and third signals and for cases in which the random queue is non-zero. As expected, the K-P model overestimates random queues. The addition of the I_a parameter to that model (Miller) improves its estimates. However, the lack of the bulk vehicle service property in both models still resulted in an overestimation of random queues. The modified Newell and Tarko-Rouphail formulas are comparable to the bulk service model.

Newell's Hypothesis

Newell recently discussed an interesting hypothesis. He suggested that from the standpoint of random queues, a signalized arterial can

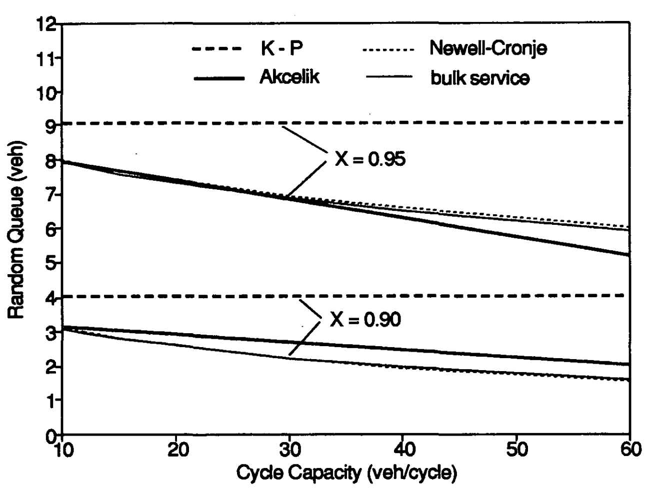

FIGURE 2 Comparison of random queue models for isolated signals.

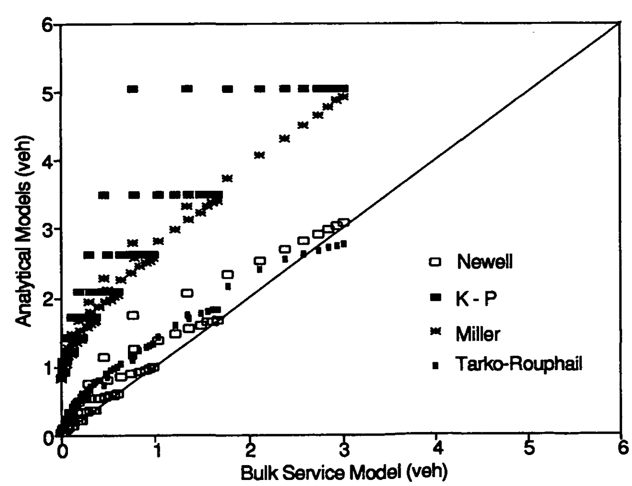

FIGURE 3 Comparison of random queue models for signalized networks.

be very easily considered as one system (*12*). He went on to state that the total random queue (and random delay) along *all* arterial signals is equivalent to the random queue that would be observed at the critical intersection were that intersection operating in isolation. The (limiting) assumption is made that there are no turning movements along all subject intersections. Approximate random delay formulas were also developed when turning movements are present, but only for modest levels.

Newell's hypothesis was tested using the same system of three signals described earlier. Figure 4 depicts the total random queue at the three signals as a function of the individual signal capacities. The results confirmed Newell's thesis that the total random queue along a signalized arterial is much lower than the total random queue if all intersections are treated as isolated (current state of the art). The results indicate an even stronger reduction than that hypothesized by Newell. It appears that Newell's estimate should be considered as an upper bound for total random queue, at least in cases in which turning movements are negligible. For practical purposes, however, his hypothesis provides much better random queue estimates than most of the formulas cited earlier.

ILLUSTRATIVE EXAMPLES

Two examples are provided to illustrate the model sensitivity to key traffic parameters and to highlight one or more stream transformations described earlier. In the first example, a single traffic stream is examined between two intersections (Figure 5, *top*). The arrival distribution at Intersection 1 is described by a Poisson process. Filtering at Intersection 1 causes a significant reduction in the variability of the departure process (Figure 5, *middle*). The resulting random queue distribution at the second signal is compared with the distribution when the upstream signal does not exist (Figure 5, *bottom*). In this case, the expected random queue at Intersection 2 is less than a third of the value computed assuming no filtering (2.60 versus 7.95). Furthermore, the total expected random queue in the system (1.82 + 2.60 = 4.42) falls far short of the expected queue length at Signal 2, assuming random arrivals (7.95). This confirms the results shown in Figure 4.

In the second example the sensitivity of the expected random queue at a downstream intersection to midblock flow levels is inves-

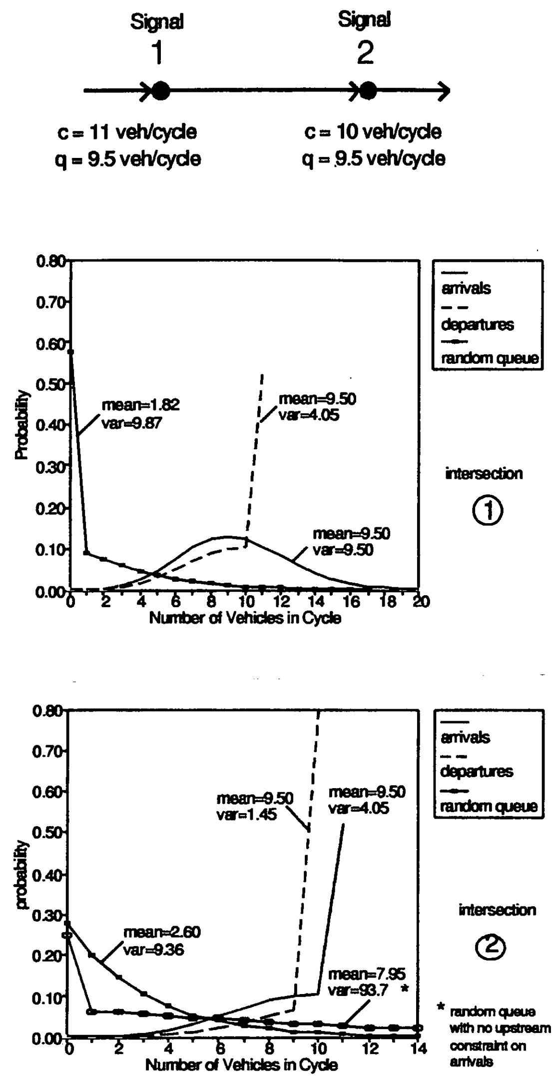

FIGURE 5 Single stream example.

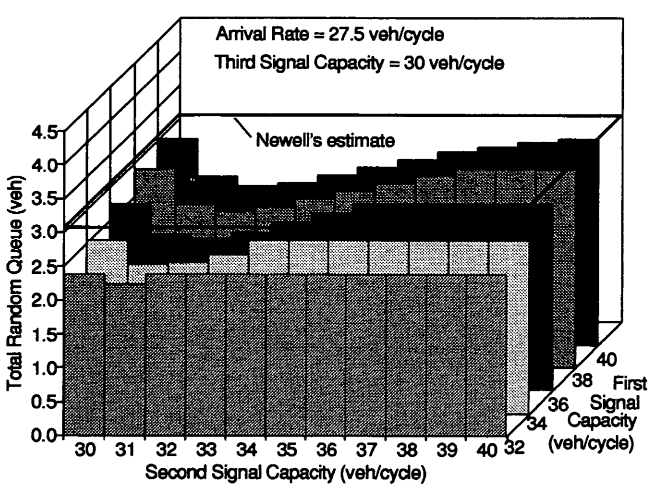

FIGURE 4 Newell's hypothesis evaluation.

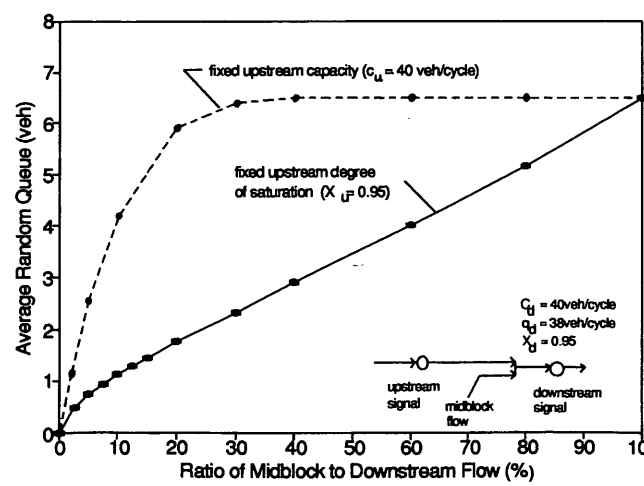

FIGURE 6 Midblock flow effect on random queue.

tigated (Figure 6). Here, stream merging and filtering effects are examined. Downstream conditions are kept fixed, including signal capacity, flow, and degree of saturation. For the upstream conditions, two scenarios are analyzed. In the first (dotted line in Figure 6), the upstream signal capacity is kept fixed while the midblock flow contribution to the total flow is allowed to increase. Consider the point at which the ratio of midblock to total flow is 30 percent. Here, the random queue reaches its maximum value. The upstream signal contribution is 70 percent of the total flow, or 26.6 veh/cycle, and its capacity is 40 veh/cycle, yielding a degree of saturation equal to 0.67. Consequently, the departure distribution is virtually unaffected by the capacity constraint (i.e., negligible filtering) and can be reasonably approximated by a Poisson process. Obviously, the combination of two Poisson processes (from signal and midblock) also produces a Poisson process with an equilibrium queue equivalent to the maximum value indicated in Figure 6.

In the second scenario (solid line), the midblock flow contribution to total flow is also increased, but the degree of saturation at the upstream intersection is maintained as fixed (0.95): for example, emulating the operation of an actuated or adaptive controller operation. Consequently, upstream filtering is active at all flow levels. The expected random queue at the downstream intersection varies almost proportionally to the portion of midblock flow in the total stream. The results clearly demonstrate the significance of midblock and turning (i.e., unfiltered) flows on random queue estimates and justify further research to incorporate that effect into analytical models of random delays and queues for signalized networks.

CONCLUSIONS

A general-arrival, bulk service time queueing model has been formulated for the study of random queues in signalized networks. The model is predicated on the occurrence of three traffic stream transformations in the network: merging, splitting, and filtering. The model is applied to steady-state conditions (traffic intensity < 1.0) but can be eventually converted to a time-dependent form to account for the effects of oversaturation. The study yielded the following conclusions:

1. Models for random queues (or delay) that are based on the Poisson arrival process (e.g., isolated intersections) are not generally transferable to networks because of the filtering effect of upstream signals.
2. Filtering tends to reduce the size of random queues. Although this finding is consistent with earlier observations by Newell (*12*), the observed reductions were even higher than Newell's estimates.
3. When two or more traffic streams merge, the resulting downstream random queue is dependent on the level of filtering that has taken place before merging. If streams are unfiltered (e.g., midblock flows, or signal departures at very low volume-to-capacity ratios), the random queue will be similar to that expected at an isolated intersection. Highly filtered streams, however, can substantially reduce random queues.
4. There is a need to consider incorporating the model results into network signal timing software, since many control strategies use a minimum delay or queue criterion for signal optimization. If methods for estimating random queues in networks must be revisited, then signal strategies that rely on such estimates should be examined.

REFERENCES

1. *Special Report 209: Highway Capacity Manual.* TRB, National Research Council, Washington, D.C., 1985.
2. Newell, G. F. Approximation Methods for Queues with Application to the Fixed-Cycle Traffic Light. *SIAM Review*, Vol. 7, 1965.
3. Webster, F. V. *Traffic Signal Settings.* Road Research Laboratory Technical Paper 39. Her Majesty's Stationery Office, London, England, 1958.
4. Haight, F. A. *Mathematical Theories of Traffic Flow.* Academic Press, New York, 1963.
5. Darroch, J. N. On the Traffic-Light Queue. *Annals of Mathematical Statistics*, No. 35, 1964, pp. 380–388.
6. Gazis, D. C. *Traffic Science.* Wiley-Interscience, 1974, pp. 148–151.
7. McNeil, D. R. A Solution to the Fixed-Cycle Traffic Light Problem for Compound Poisson Arrivals. *Journal of Applied Probability*, No. 5, 1968, pp. 624–635.
8. Van As, S. C. Overflow Delay at Signalized Networks. *Transportation Research*, Vol. 25A, No. 1, 1991, pp. 1–7.
9. Kimber, R., and E. Hollis. *Traffic Queues and Delays at Road Junctions.* Report 909. U.K. Transport and Road Research Laboratory, Crowthorne, Berkshire, England, 1979.
10. Tarko, A., N. Rouphail, and R. Akçelik. Overflow Delay at a Signalized Intersection Approach Influenced by an Upstream Signal: An Analytical Investigation. In *Transportation Research Record 1398*, TRB, National Research Council, Washington, D.C., 1993, pp. 82–89.
11. Akçelik, R. *Traffic Signals: Capacity and Timing Analysis.* Research Report 123. Australian Road Research Board, Nunawading, 1981.
12. Newell, G. F. Stochastic Delays on Signalized Arterial Highways. *Proc., 11th International Symposium on Transportation and Traffic Flow Theory*, Elsevier, New York, 1990, pp. 589–598.
13. Miller, A. J. *Australian Road Capacity Guide—Provisional Introduction and Signalized Intersections.* Bulletin 4. Australian Road Research Board, Nunawading, 1968 (superseded by Research Report 123, Australian Road Research Board, 1981).
14. Olszewski, P. Modelling of Queue Probability Distribution at Traffic Signals. *Proc., 11th International Symposium on Transportation and Traffic Flow Theory*, Elsevier, New York, 1990, pp. 569–588.
15. Cronje, W. B. Analysis of Existing Formulas for Delay, Overflow, and Stops. In *Transportation Research Record 905*, TRB, National Research Council, Washington, D.C., 1983, pp. 89–93.

Publication of this paper sponsored by Committee on Traffic Flow Theory and Characteristics.