ON THE TEMPORAL DISTRIBUTION OF PEAK TRAFFIC DEMANDS: A MODEL AND ITS CALIBRATION

A.S. Alfa
Department of Civil Engineering
University of Manitoba

J.A. Black
Department of Transport Engineering
School of Civil Engineering
University of New South Wales

W.R. Blunden
University of New South Wales

ABSTRACT:

The paper reports on research that developed a rationale for the mechanism of peak spreading of traffic and validated the model experimentally. As details of the individual optimising model have been published elsewhere, the focus here is on defining a bottleneck situation on the transport network to articulate the theory and on conducting sensitivity analyses to simulate the effects of traffic management schemes. Peak-hour traffic data for Sydney's major bottleneck, the Sydney Harbour Bridge, are used to calibrate the cost coefficients of transport impedance that include penalties for being early, being late and being delayed in the traffic stream. Implications for current transport engineering and planning practice are suggested.

Or in the night, imagining some fear,
How easy is a bush suppos'd a bear!

(Midsummer-Night's Dream, Act V, Scene I)

INTRODUCTION

The fear of road traffic congestion is one important driving force behind freeway and expressway proposals for new urban areas. Although the estimation of probable peak-hour flows on a given facility is a complex task, transport engineering practice relates the intensity of this traffic demand to a fixed-time period by applying empirical rules of thumb (peak-hour ratios) that calculate design-hour volumes as a proportion of total daily traffic. Not only does this lead to arbitrary facility design in terms of the number of traffic lanes required, as shown by Shallal and Khan (1980, Table I, p.76), for instance, but it ignores the mechanism of peak spreading. Without an understanding of this temporal distribution of traffic, it is easy to see why a bush becomes a bear (see, Blunden, 1982). This neglect of the temporal distribution of traffic is a problem of practical consequence because a demand rate must be specified for traffic assignment purposes.

Although we have an empirical picture of the temporal distribution of road traffic during a rail strike (Clunas, 1984), research was directed towards developing a rationale for the mechanism of peak spreading of traffic and validating the model experimentally. Intuitively, a given demand on a transport network must spread itself over a finite time - a great deal of data have been compiled on the extent of the rush-hour duration and its peaking characteristics in cities of different size (Peat, Marwick, Mitchell and Co, 1972). Also, planning emphasis on making better use of existing transport facilities instead of constructing new ones (Remak and Rosenbloom, 1976, 1979; Rosenbloom, 1978) involves the alteration of work schedules as an important policy instrument in alleviating congestion by flattening the peak-travel demands (Julian, 1971; O'Malley and Selinger, 1973; Transportation Research Board, 1980). In predicting the effect of such changes in work schedules on traffic congestion it is helpful to have a theoretical understanding of how the peak demand develops and the phenomenon of peak spreading.

The model proposed takes, as a starting point, the proposition that commuters seek to minimise 'transport impedance' and in doing so may vary their starting times for a journey-to-work. Unlike the conventional behavioural approach that represents transport impedance by the generalised cost of travel, this approach assumes that travellers attach 'cost coefficients' to being early, to being late, and to delays in the traffic stream. Therefore, it is an individual optimising model in the tradition of Wardrop (1952) and owes inspiration to the work of Gaver (1968) and Minh (1976).

A bottleneck situation on the transport network is used to articulate the model, although this simplification does not affect the generality of the theory. By specifying a target time, or a distribution of target times, at the bottleneck, the travel time from the home to the bottleneck is ignored (or assumed to be a free-flow travel time). In effect, the distribution of the departure times from home is identical to the distribution of the arrival times at the bottleneck. Similarly, the travel time from the bottleneck to the workplace is ignored and the departure times from the bottleneck become the arrival times at the destination. A bottleneck situation with time measured in epochs is especially convenient for data collection and model calibration, as demonstrated by our case study of the Sydney Harbour Bridge.

The Sydney Harbour Bridge - 1149 m long, including approach spans - is a major bottleneck for Sydney commuters travelling across the Parramatta River and therefore is a suitable case study example for the application and calibration of our model. The steel-arch (503 m span) bridge itself is a celebrity (Spearritt, 1982) being both a household name for Sydney residents and a distinctive landmark for visitors, and having celebrated its fiftieth birthday on 19 March, 1982. The designers of the 1920s estimated the capacity of four railway lines, six road lanes and two footways as 160 trains per hour, 6,000 vehicles per hour and 40,000 pedestrians per hour. Records of the Department of Main Roads, New South Wales, for 1980 show the average annual daily flow of 158,850 vehicles with a maximum daily flow of 200,500 vehicles. With tidal flow operations, the maximum hourly flow (in the morning) was 10,850 vehicles in the direction of the major flow (southbound into the CBD) and 4,340 vehicles in the direction of the minor flow (northbound).

Before presenting the results of our investigation, the first section describes the structure of the model and defines the key parameters. This is done only in outline form in this paper: the important equations are given without discussion because the full details of the stochastic model of the temporal distribution of peak traffic demands have been explained elsewhere (Alfa and Minh, 1979). This model by Alfa and Minh can be reviewed in the broader literature on departure times, journey times and waiting times. Abkowitz (1981) and Small (1982) used logit models to study the choice of departure times but their approaches were empirical. Whereas these two models could estimate the effect of travel time on the choice of departure times they did not include how changes in the choice of departure times affect travel times. More recently De Palma (et al, 1983) also used a logit model to study the same problem. The interdependence of travel time and choice of departure times was included in their model, but they used a deterministic queueing model for estimating the waiting time. Although all these models were based on a stochastic approach, the waiting time models used were deterministic. Vickrey (1969), Hendrickson and Kocur (1981) and Fargier (1981) all followed the user-equilibrium approach to study a similar problem, but, again, deterministic queueing was employed to estimate the waiting time. (Note that Henderson (1977) later generalised Vickrey's model into a non- queueing general congestion situation.) In contrast to these approaches, the stochastic model presented by Alfa and Minh (1979) was set up as a Markov chain, and a stochastic queueing model was used to estimate the waiting time.

In this paper, the original contribution is the presentation of the results of sensitivity analyses conducted with the model: the effect on the temporal distribution of traffic by (a) changing the demand; (b) altering the capacity of the bottleneck; (c) altering the target time at the ultimate destination (staggered work hours, flexitime); and (d) assuming different cost coefficients for being early, late and delayed in the traffic. Calibration of these cost coefficients, using authentic traffic data collected in Sydney, is demonstrated. Implications for transport engineering practice, when the objective is to make more efficient use of existing resources, are explored in the concluding section.

THE MODEL

The model for the distribution of arrival times at the bottleneck is of the kind developed for headstart strategies to overcome the variability of travel times in meeting a destination target time. Although its mathematical formulation has been described fully by Alfa and Minh (1979), its specific application here to 'rush hour' traffic situations calls for a presentation of the model's definitive equations. The prime purpose of the model is to yield a limiting pattern of arrival rates (demand intensities) over a finite time - that is, a busy period - that will minimise the perceived cost to the user of a commuting (or some other repeated access) exercise.

Commuters take account of costs associated with departing from home early, arriving at work late and being delayed in the traffic over and above the 'zero-flow' travel time of their journey. By ignoring the zero-flow travel time, the arrival distribution of the input to a transport network is synonomous with the departure from home and the departure distribution with arrivals at the destination. Earliness and lateness may then be measured by the time intervals between arrival and departure and a specified time, T, the target time. The transport network is considered as a single channel queueing facility with capacity equal to that of the network, as may be determined by the maximum flow/minimum-cut theorem of Ford and Fulkerson (for example, see, Blunden and Black, 1984).

In the context of transport management and planning, an important feature of the model is that its principal input is the total demand I+1, comprising a typical commuter plus I other commuters. The arrival of the commuters is time dependent. The queueing model is established in a discrete time scale made up of equally-spaced epochs numbered 0,1,2...N, which allows us to use the discrete time queueing model developed by Minh (1977). It is assumed that all commuters arrive and depart at instants immediately prior to these epoch units. We designate the typical commuter as A, and suppose that this commuter is identical with, but acts independently of all the others. The decision making strategy is illustrated with reference to this typical commuter. In making this journey on day, d, the commuter arrives at the bottleneck at epoch, n, and experiences a delay of i (epochs) in the queue and thus would depart from the bottleneck at epoch n+i+S, where S is the service time. There are three mutually exclusive situations:

(a) $n+i+S < T$, and A arrives at the destination early by an amount $T-n-i-S$, and attaches a cost $C_e(n,i)$ for so doing;

(b) $n+i+S > T$, and A arrives late by $n+i+S-T$ time units and so incurs a cost $C_\ell(n,i)$; and

(c) $n+i+S = T$; A arrives on time.

In addition, A attaches a cost $C_\omega(i)$ for delay in the traffic, a factor not considered by Gaver (1968).

The total perceived cost of A's commuting circumstances, C(n,i) is given by:

$$C(n,i) = C_\omega(i) + \begin{cases} C_e(n,i) & \text{if } n+i+S < T \\ 0 & \text{if } n+i+S = T \\ C_\ell(n,i) & \text{if } n+i+S > T. \end{cases}$$

Commuter A now decides to 'experiment' by changing the time of arrival at the bottleneck (or, in other words, departure from home) to epoch, m, hoping that the consequent delay, j, will result in a total perceived cost, C(m,j) that results in {C(n,i) - C(m,j)} being greater than or equal to zero. If the delay is not reduced, a different arrival time is chosen. This benefit measure is designated as:

$$[C(n,i) - C(m,j)]^+$$

where $[\ldots]^+$ indicates it is taken into account in the model only when positive.

How A decides on a value for m that will improve the journey for the following day is described in Alfa and Minh (1979). It will suffice here to define $W_i^n(d)$ as the probability that A experiences a delay of i, given arrival at epoch, n, on day, d. The anticipated reduction in cost, $q_{n,m}(d)$ by changing to epoch, m, on day, d+1, may be written:

$$q_{n,m}(d) = \sum_{j=0}^{I\times S} \sum_{i=0}^{I\times S} [C(n,i) - C(m,j)]^+ \ W_i^n(d) \ W_j^m(d+1).$$

Commuter A must estimate W_j^m (d+1) and the only plausible basis for doing this is to assume that the other I commuters do not change their arrival process for the following day and so the delay distribution is the same as day, d. We write $\tilde{q}_{n,m}(d)$ as a meaningful estimate by A of $q_{n,m}(d)$, and so

$$\tilde{q}_{n,m}(d) = \sum_{j=0}^{I\times S} \sum_{i=0}^{I\times S} [C(n,i) - C(m,j)]^+ \ W_i^n(d) \ W_j^m(d)$$

where, $W_j^m(d)$ is A's estimate of $W_j^m(d+1)$.

On the basis of these estimates, A considers a benefit is derived from changing from epoch n to m if $\tilde{q}_{n,m}(d) > 0$ (Alfa and Minh, 1979, p.321). From now on A can lose identity and become just one other commuter with a transition probability, $t_{n,m}(d)$, of changing to epoch, m, on day, (d+1), given that the commuter arrived at epoch, n, on day d:

$$t_{n,m}(d) = \tilde{q}_{n,m}(d) \Big/ \sum_{m=1}^{N} \tilde{q}_{n,m}(d).$$

Let T(d) be an NxN transition matrix such that

$$[T(d)]_{n,m} = t_{n,m}(d).$$

If we now let the distribution of arrivals on day, d, be:

$$\underline{\Pi}(d) = [\pi_1(d), \pi_2(d) \ldots \pi_N(d)]$$

then a commuter's arrival time on day (d+1) may be set up as a classical Markov chain:

$$\underline{\Pi}(d+1) = \underline{\Pi}(d) \times T(d).$$

The sensitivity of $\lim_{d\to\infty} \underline{\Pi}(d)$ to the main parameters of the model is studied in the next section but before doing so we present in Table 1 the results of a typical iteration path from an initial distribution of arrivals that is assumed to be uniform. In this worked example, N = 10, I+1 = 5, T = 8 and the relative cost coefficients are C_e = 1, C_ℓ = 2 and C_ω = 4. The smooth and rapid convergence of the model provides a measure of validation of its 'mathematical' structure.

Table 1. A Typical Path of Iteration of Arrival Probabilities by Epochs

Epoch	Arrival Probability, Π_n, at Each Iteration									
n	Start	1st	2nd	3rd	4th	5th	6th	7th	8th	9th
1	0.1000	0.0257	0.0408	0.0337	0.0360	0.0353	0.0355	0.0355	0.0355	0.0355
2	0.1000	0.0405	0.0579	0.0507	0.0531	0.0524	0.0527	0.0526	0.0526	0.0526
3	0.1000	0.0620	0.0815	0.0743	0.0765	0.0769	0.0760	0.0760	0.0760	0.0760
4	0.1000	0.0943	0.1140	0.1064	0.1086	0.1082	0.1082	0.1082	0.1082	0.1082
5	0.1000	0.1348	0.1490	0.1411	0.1437	0.1433	0.1433	0.1433	0.1433	0.1433
6	0.1000	0.1905	0.1839	0.1793	0.1825	0.1814	0.1816	0.1816	0.1816	0.1816
7	0.1000	0.2443	0.1757	0.1967	0.1927	0.1925	0.1930	0.1928	0.1929	0.1928
8	0.1000	0.1268	0.0960	0.1195	0.1094	0.1125	0.1118	0.1118	0.1119	0.1119
9	0.1000	0.0584	0.0631	0.0667	0.0638	0.0652	0.0647	0.0648	0.0648	0.0648
10	0.1000	0.0228	0.0381	0.0319	0.0337	0.0333	0.0333	0.0333	0.0333	0.0333

(Note: N=10; I+1=5; T=8; C_e=1; C_ℓ=2; and C_ω=4)

SENSITIVITY ANALYSIS OF MODEL PARAMETERS

In the conceptual model the parameters that determine the pattern of the temporal distribution of peak traffic demand are: the total population of commuters; the capacity of the transport element; the time that commuters wish to arrive at their destination; and the perceived cost that commuters attach to earliness, lateness and delays in the traffic. This section examines how each one affects the temporal distribution of peak demand, how some of them can be controlled to reduce traffic congestion, and shows the method to estimate the cost coefficients for each situation. The number of commuters and the saturation flow rates are known but the target time is known only implicitly and the cost factors are not known and require calibration.

The Number of Commuters

Results from the model show that when the total number of commuters increases the number of arrivals at each epoch increases at the bottleneck but that the percentage increase in the number of arrivals at each epoch is greater during those epochs with a lower initial demand rate. The mode of the peak demand tends to shift slightly to an earlier time epoch, as demonstrated by the following worked example.

Consider a system with a capacity of 200 commuters per unit epoch over the period N = 15 and let T = 10. Suppose all the users of this system attach linear costs with the rates \$1.00, \$2.00, and \$3.00 per unit time to earliness, lateness and delays in the traffic, respectively. For populations of 400, 800 and 1200 commuters, Figure 1 shows the temporal distribution. This spreading of the demand rate as the total demand increases confirms the intuitively obvious notion of a self-regulating system.

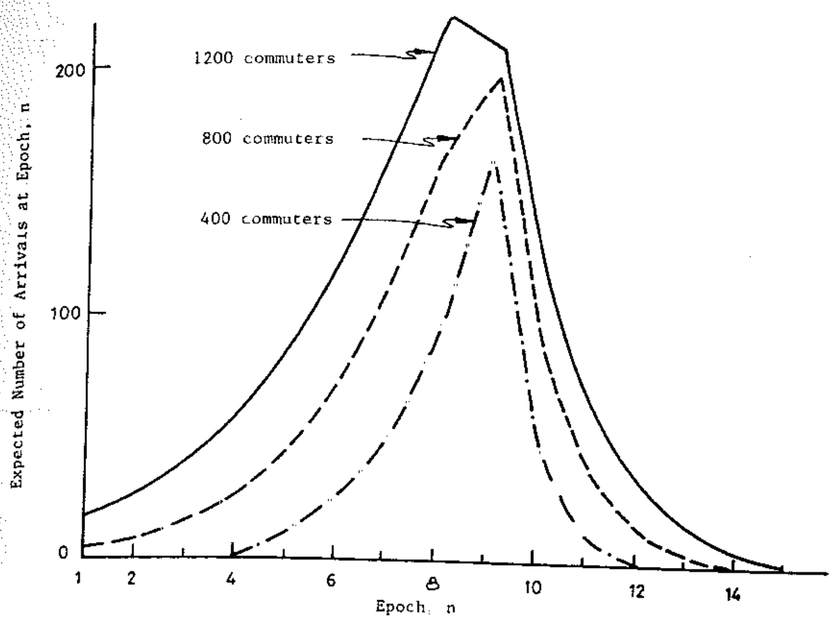

Figure 1. The Effect of Changing the Total Number of Commuters on the Temporal Distribution of Peak Traffic Demand

Effect of Change Capacity

Again, for illustrative purposes, suppose a bottleneck has a capacity of 100 vehicles per minute and that 6000 commuters all wish to depart the bottleneck at 8.30 am in order to arrive at their destination at the desired time. The system is observed for two hours between 7.10 am and 9.10 am. The spacing of the epochs is 10 minutes, so N = 12 and a satisfactory result for the commuter is departing the bottleneck at either epoch $T_1 = 8$ or $T_2 = 9$. If we further assume that all the commuters attach linear costs of \$10.00, \$20.00 and \$5.00 per unit time to earliness, lateness and delays in the traffic, respectively, then the resulting temporal distribution of traffic demand is shown in Figure 2.

If the capacity of the bottleneck is increased to 120, to 150, to 200 and then to 300 vehicles per minute, the resulting four distributions of traffic demand are calculated and are shown in Figure 2. An increase in capacity encourages a further intensification of the peak demand rate, although the duration of the total demand is reduced.

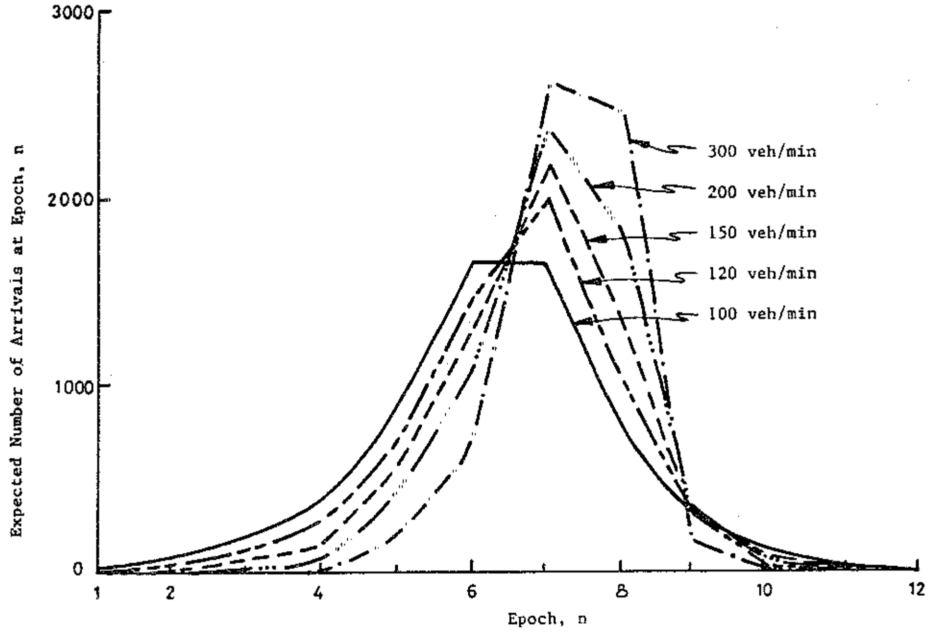

Figure 2. The Effect of Changing the Capacity of the Bottleneck on the Temporal Distribution of Peak Traffic Demand

The Target Time - Staggering, Flexible Working Hours

The effect of the target time on the temporal distribution of traffic demand is obvious because any shift in the position of that target time in a real time axis merely shifts the whole temporal distribution of traffic demand by the same amount and in the same direction provided that the time range, N, is defined to be long enough. However, its effect on the position of the mode of the peak of the traffic demand depends on the value of the cost parameters. Generally, this mode occurs at some epochs before the target time: the higher the cost for late arrivals then the more will this mode move to earlier epochs. The reverse happens when the cost for early arrivals increases.

Let us define two time epochs T_1 and T_2 between which commuters are free to start work. The total cost incurred by a commuter under this scheme would be:

$$C(n,i) = C_\omega(i) + \begin{cases} C_e(n,i) & ; \text{ if } n+i+S < T_1 \\ 0 & ; \text{ if } T_1 \leqslant n+i+S \leqslant T_2 \\ C_\ell(n,i) & ; \text{ if } n+i+S > T_2 . \end{cases}$$

In order to introduce flexible working hours into the analysis the cost structure has been modified. Intuitively, we would expect that flexible work hours should result in the peak demand being more spread out, provided the cost attached to delay is not zero. The time that the mode of the peak demand occurs will be shifted to the left or to the right depending on whether the cost attached to lateness is greater than that attached to earliness or vice versa.

The base case is a system where 4000 commuters wish to depart from a bottleneck at 8.30 am in order to arrive at work on time. The saturation flow at the bottleneck is 50 vehicles per minute and $N = 20$ and $T = 15$. Instead of having this target time $T = 15$, flexitime can be introduced by taking a range between $T_1 = 13$ and $T_2 = 17$. Figure 3 shows the arrival distribution at the bottleneck for the base case and for flexible working hours. The effect of flexible work hours seems to be beneficial in terms of reducing the peak, at least for the example considered here (the value of the cost parameters influences the extent to which this strategy alleviates traffic congestion). When the cost attached to delay is extremely large compared to other costs, little difference between both cases would be found.

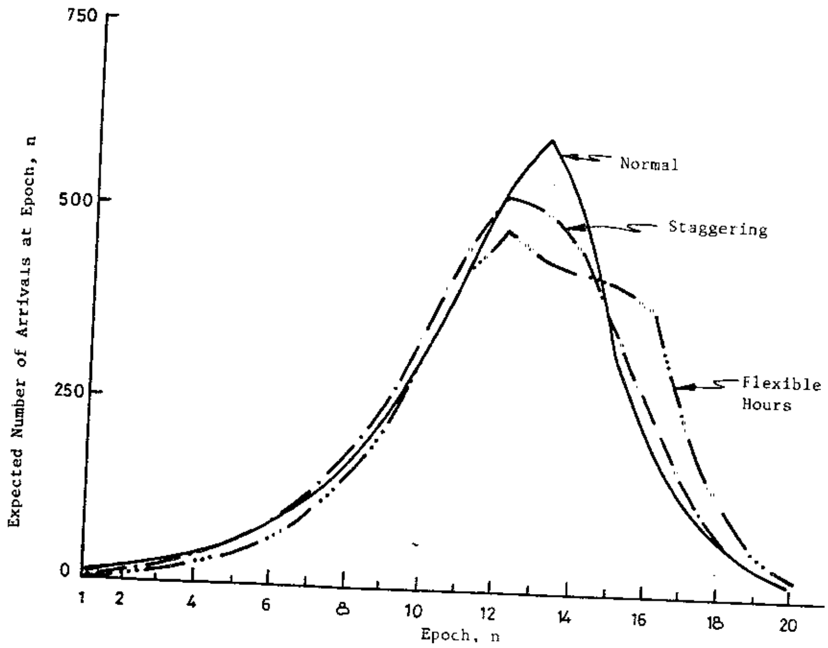

Figure 3. The Effect of Variable Working Hours on the Temporal Distribution of Peak Traffic Demand

The cost structure is redefined to introduce staggered work hours into the analysis and the commuters are classified into groups who are required to start work at some specified time. Let the number of such groups for a particular system be G and define T^g to be the target time of group, g, ($1 \leq g \leq G$). The proportion of commuters in group g is p_g, such that $\sum_{g=1}^{G} p_g = 1$, and let $C(n,i)g$ be the total cost to a commuter of group g type, given arrival at the bottleneck at epoch, n, with delay of i units of time, such that:

$$C(n,i)g = C_\omega(i) + \begin{cases} C_e(n,i) & ; \text{ if } n+i+S < T^g \\ 0 & ; \text{ if } n+i+S = T^g \\ C_\ell(n,i) & ; \text{ if } n+i+S > T^g . \end{cases}$$

Consider the typical commuter A once more: the probability that A belongs to group g is given by p_g. Therefore, given that the commuter arrived at the bottleneck at epoch, n, and was delayed i units of time, then the total cost incurred, C(n,i), is given by:

$$C(n,i) = \sum_{g=1}^{G} p_g \times C(n,i)g.$$

By assuming that each one of the commuters in the G group behave in a similar manner as A, the problem can then be formulated as in the case of a single target time.

The base case is defined as before but instead of having one target time, let us have G = 3, $T^1 = 13$, $T^2 = 15$ and $T^3 = 17$ and also let $p_1 = p_3 = 1/4$ and $p_2 = 1/2$. Figure 3 shows the results of the analysis: staggered working hours relieve the congestion problem during the peak period, although not by as much as in the case of flexible work.

Commuters having different destinations situated after the bottleneck can be also approximated by the same model as for staggered working hours, provided that queueing does not occur after the commuters depart the bottleneck. By grouping those commuters with a common destination, and by knowing both the time they wish to arrive at their destination and the remaining travel time after they depart the bottleneck, the time they would like to depart the bottleneck can be obtained.

Cost Coefficients

It is plausible, though worthy of verification, that commuters attach different perceived costs to earliness, lateness and delays. However, as we shall be discussing the aggregate values of these coefficients for all the population using a particular system, commuters are all assumed to have identical values. The magnitude and functional forms of these values affect the temporal distribution of the peak demand in terms of the positioning of the mode of the peak and the spreading mechanism of the demand. There are only two things we know for certain about these cost coefficients: one, the costs to earliness, lateness and delay increase monotonically with the amount of time a commuter is early, late and delayed, respectively; and two, that $C_e(n, T-n-S) = 0$, $C_\ell(n, T-n-S) = 0$ and $C_\omega(0) = 0$; for $n+S \leqslant T$ and $n,S,T \geqslant 0$. Provided that all these conditions are satisfied, the model is able to reproduce the general pattern of the temporal distribution of the peak traffic demand noticed in most traffic systems, irrespective of the magnitudes or functional forms assumed for the cost parameters.

For example, consider the base case in the previous sub-section and suppose that instead of the costs being linear they are proportional to the square of the amount of time the commuter is early, late or delayed, respectively. That is:

$$C_e(n,i) = C_e \times (T-n-i-S)^2$$
$$C_\ell(n,i) = C_\ell \times (n+i+S-T)^2 \text{ ; and}$$
$$C_\omega(i) = C_\omega \times (i)^2.$$

The resulting arrival time distribution from this is compared to the one with linear costs in Figure 4. The same pattern of traffic is produced with a slight variation to the spread and the position of the mode of the peak.

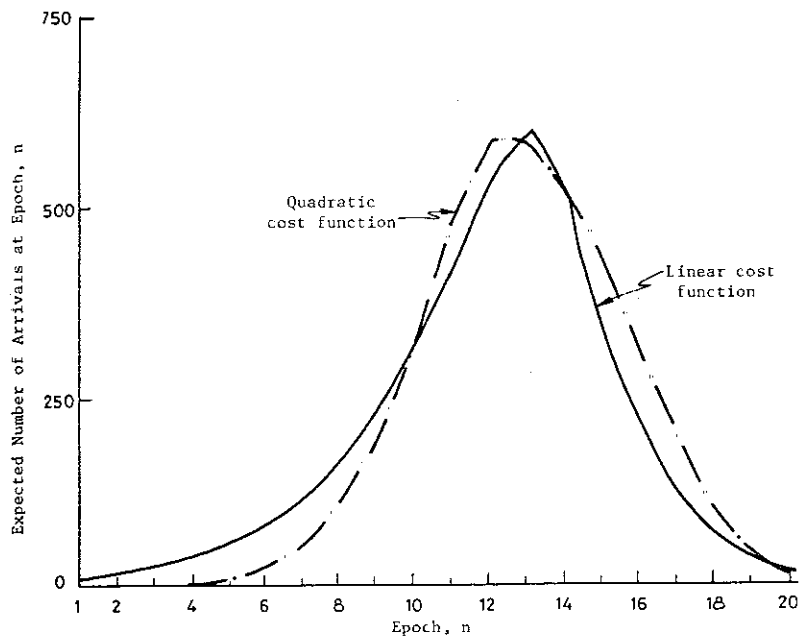

Figure 4. The Effect of the Functional Form of the Cost Parameters on the Temporal Distribution of Peak Traffic Demand

The magnitudes of these cost coefficients, on the other hand, do affect significantly the position of the mode of the peak and the spread of the demand. If the cost that the commuters attach to each unit time of delay increases, it is natural that they will try to avoid long delays more often and thus further spread their demand over the time so as to reduce their delay. In addition, if the cost that they attach to lateness increases they will prefer to arrive at their destinations earlier than run the risk of being late. Hence, the whole distribution of the temporal traffic demand will be shifted to an earlier time (i.e. shifted to the left). Conversely, if the cost to earliness is increased, the distribution would shift further to the right.

As previously, the base case, a, is with C_e = \$1, C_ℓ = \$2 and C_ω = \$3. Consider example b, in which the cost for delay is increased to \$50.00 per unit time, and otherwise let everything else remain as in a. Consider examples c and d, in which the cost for lateness is increased to \$50.00 per unit time and the cost for earliness is increased to \$50.00 per unit time, respectively. Figure 5 gives the resulting arrival distributions that show a marked difference in shape for the four cost assumptions. The implication of this is that knowledge of the functional forms, and in particular the magnitudes of these cost parameters for any peak traffic situation, is imperative if our model is to be used for predictive purposes. A study of this problem is reported in the next section.

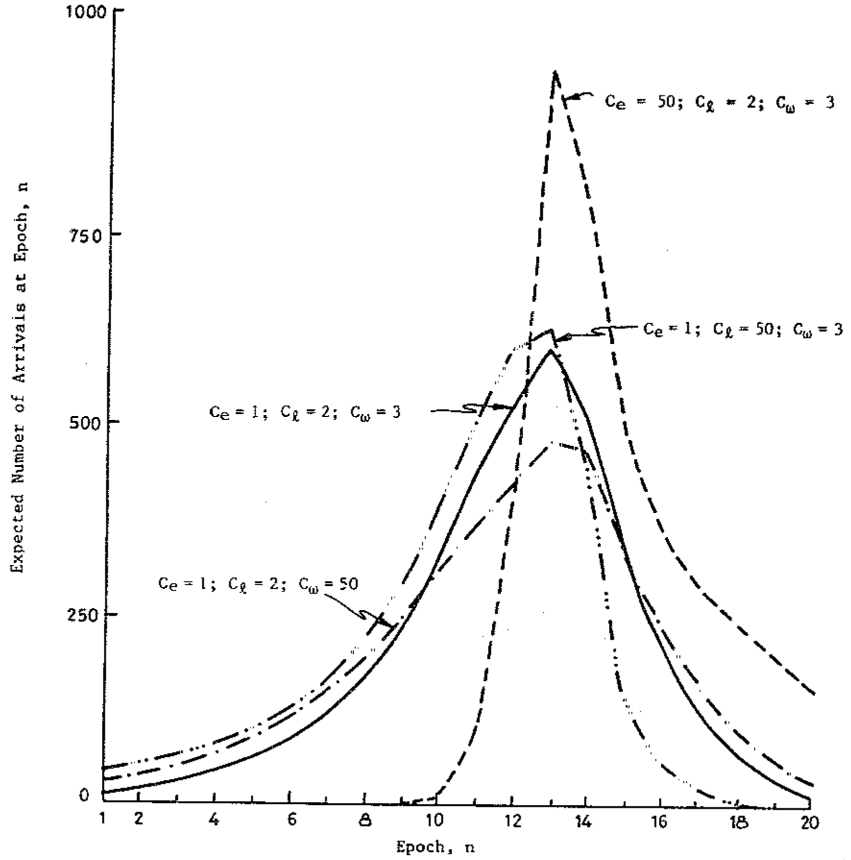

Figure 5. The Effect of Changing the Cost Parameters on the Temporal Distribution of Traffic Demand

ESTIMATION OF THE COST PARAMETERS

Model Calibration - Procedure

Much could be said about the functional form of the cost parameters (Alfa, 1979, pp. 87-91) but here we merely state that:

$$C(n,i) = C_\omega(i) + \begin{cases} C_e \times (T-n-i-S) & ; \text{ for } n+i+S \leq T \\ C_\ell \times (n+i+S-T) & ; \text{ for } n+i+S \geq T. \end{cases}$$

The parameter estimation was carried out using the method of least squares. The method assumes that there is an arrival probability vector $\hat{\underline{\Pi}}$ estimated from the arrival distribution observed at the particular major bottleneck. Also, for any given values of C_e, C_ℓ and C_ω, there exists a probability vector $\underline{\Pi}$ that can be evaluated for that particular bottleneck using $\underline{\Pi} = \underline{\Pi} \times \underline{T}$. The estimated values of these cost parameters that describe the traffic at this bottleneck are those values that minimise U, the sum of the squares of the differences, between $\hat{\underline{\Pi}}$ and $\underline{\Pi}$, where:

$$U = (\Pi - \hat{\underline{\Pi}}) \times (\underline{\Pi}-\hat{\underline{\Pi}})^T$$

and $(\ )^T$ denotes the transpose of the matrix in the bracket. Minimising U can only be achieved by a numerical method, which requires considerable computation time.

However, the number of iterations, and hence the computation time required, was reduced by eliminating one of the unknown variables. As we are only interested in the relative trade-offs between the cost coefficients, knowing the values of two of these coefficients relative to the third provides us with sufficient information. Therefore, let $C_e = 1$ and define the costs C_ℓ and C_ω in terms of C_e with the assumption that $C_e > 0$ in real life. Although this modification will change the values of $\tilde{q}_{n,m}$ it will not affect the values of $t_{n,m}$ and does not change the structure of the problem. The calibration problem is an unconstrained, bivariate non-linear optimisation problem (Alfa, 1982). The partial derivatives of the objective function, U, with respect to these two variables C_ℓ and C_ω, cannot be obtained analytically. The method developed by Powell (1964), which does not require calculating the function's derivatives, was used - the computer program for this method is available in Kuester and Mize (1973).

Estimation of the Cost Parameters, Sydney Harbour Bridge

All of the travellers using the Sydney Harbour Bridge during the peak period are assumed to be commuters. Collecting the data of arrivals at a bottleneck where the queues that develop can grow to a great length requires a lot of field workers situated at different points before the bottleneck in order to record arrival counts at the end of queue that keeps fluctuating in its position. The data were collected in 1977 on the Cahill Expressway as it approaches the bridge entrance. Both the number of vehicles that arrive at and depart from this bottleneck were recorded from the time the peak period queue began to build up at 7.30 am up to the time it dissipated at 9.00 am. The departure rate was used to compute the capacity (saturation flow) of the lanes at the bottleneck. The capacity of this section of the expressway, at the bridge, is approximately 65 vehicles per minute and the total demand during this peak period was about 5,753 vehicles.

By defining epoch spacings as 10 minutes then N = 9. Although southbound traffic has different destinations, they were grouped into four distinct geographical zones as shown by Figure 6. The average travel time between the adjacent zone centroids is assumed to be 10 minutes. A commuter with a destination in zone g (g = 1,2,3,4) belongs to group g. If every commuter must reach their respective destinations at 9.00 am then those whose destination is in zone 1 will want to depart the Harbour Bridge at 8.50 am and those for zones 2, 3 and 4 at 8.40 am, 8.30 am and 8.20 am, respectively.

If C(n,i)g is the total cost to a commuter of group, g, then:

$$C(n,i) = \sum_{g=1}^{4} p_g \times C(n,i)g$$

where, p_g is the ratio of those belonging to group g. A morning peak origin-destination traffic study in 1974 (Clarke Gazzard Voorhees, 1974) gave, for the zoning system in Figure 6, values of $p_1 = 0.1099$; $p_2 = 0.2569$; $p_3 = 0.4766$; and $p_4 = 0.1566$.

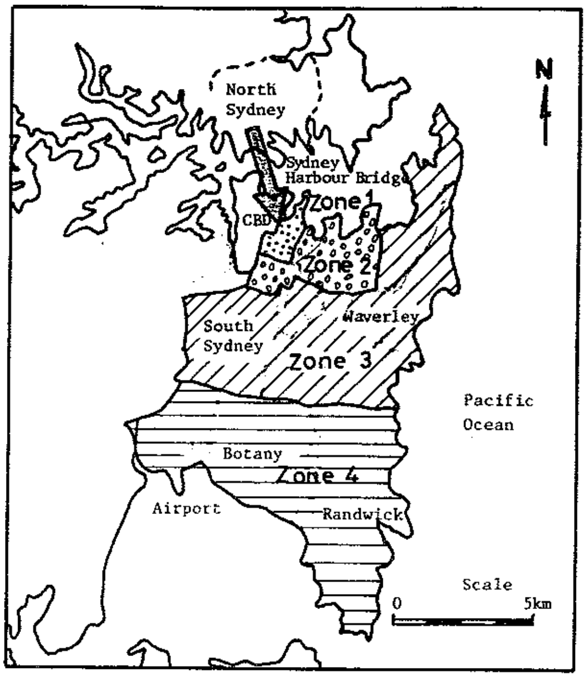

Figure 6. The Destination Zones of the Southbound Sydney Harbour Bridge Traffic Travelling on the Cahill Expressway during the Morning Peak Period, 1977.

The calibration of the model gave the following estimates:

$$\tilde{C}_{\ell} = 1.35; \quad \tilde{C}_{\omega} = 7.5; \quad \text{and} \quad U = 0.000358.$$

The aggregate cost attached to lateness is about 1.35 of that attached to earliness for the commuters employed south of the bridge and residing north of it. However, they attach about seven and a half times as much to delay in the traffic. The fact that a commuter's lateness or earliness to work depends on his ability to predict the anticipated delay, gives us ground to suspect that the cost attached to delay is very much higher, and more dominant, than either one of the other costs. The least squares curve of the arrival probability $\underline{\Pi}$, estimated using $\tilde{C}_{\ell}$ and $\tilde{C}_{\omega}$ is shown in Figure 7a where it is compared with actual traffic data.

The data for the northbound traffic were collected a year later, but in exactly the same manner. The saturation flow of the bottleneck studied was about 28 vehicles per minute and the total demand during this peak-period was about 2800 vehicles. Much of this northbound traffic is travelling to North Sydney, which is an extension of the Sydney CBD on the north shore of the Parramatta River. Here, one destination zone is assumed where the travel time within the zone does not exceed 10 minutes. Again, if it is assumed that each commuter must reach their destination at 9.00 am and it takes 10 minutes to travel from the bottleneck exit to this destination, they have to depart the bottleneck by 8.50 am.

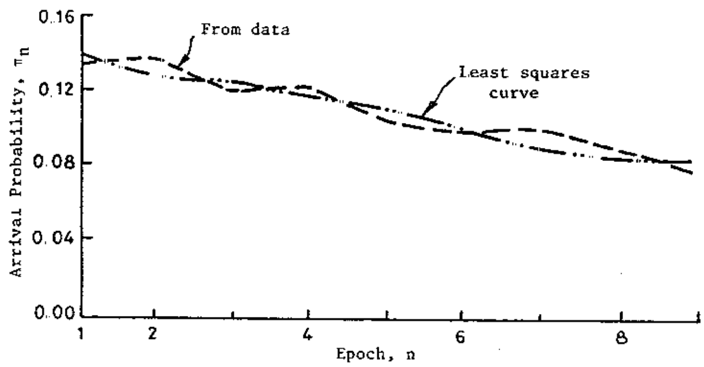

Figure 7a: Arrival Probabilities for the Southbound Traffic Sydney Harbour Bridge 1977

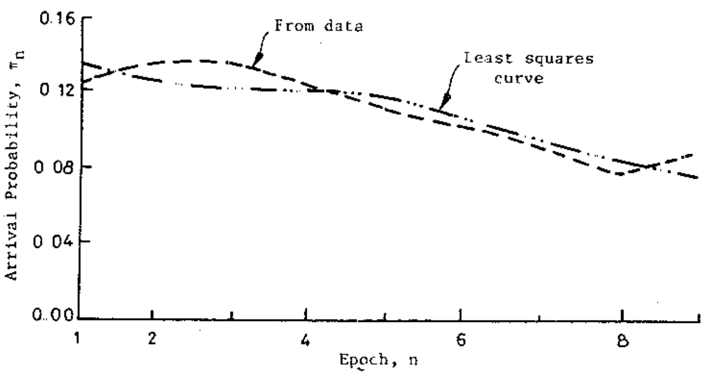

Figure 7b: Arrival Probabilities for the Northbound Traffic Sydney Harbour Bridge, 1978

Figure 7. Arrival Probabilities for (a) the Southbound Traffic on the Sydney Harbour Bridge, 1977 and (b) the Northbound Traffic on the Sydney Harbour Bridge, 1978.

The calibration of the model gave the following estimates for northbound traffic:

$$\tilde{C}_\ell = 1.6; \quad \tilde{C}_\omega = 6.0; \quad \text{and} \quad U = 0.000729.$$

The least-squares curve of the arrival probability, $\underline{\Pi}$, is plotted in Figure 7b. The relatively high values of C_ω show that although commuters do not wish to arrive at work late or too early, the most undesirable aspect of the travel is excessive delay in the traffic stream. This leads to the conclusion that most commuters would rather arrive at work early than be late or be delayed for a long time in the system, because $\tilde{C}_\ell > 1$ and $\tilde{C}_\omega$ (> 1) is quite large. In fact, some of them, if necessary, would settle for a bit of lateness if by doing so they could reduce, by a considerable amount, the excessive delay they would encounter, because $C_\omega/C_\ell > 1$. This supports what is intuitively obvious and well documented in peak-hour traffic studies cited in the introduction: as the total demand increases it will spread itself out temporally rather than intensifying further.

CONCLUSION

The results obtained on the sensitivity of the commuting exercise to total demand level, transport capacity, target time arrangements, and perception of costs confirm the every-day experience of road users and traffic management professionals alike. In the area of planning, the underpinning of the self-adjusting nature of the demand rate pattern by the theoretical model is of special importance, notably the relatively high value of perceived cost of excess delay in the traffic stream itself. The importance of the extra delay over and above the 'free flow' travel time suggests that unwanted delay is a force that spreads the peak - even without formal flexitime or staggered work hour arrangements.

This finding has implications for the behavioural specification of the generalised cost of travel by private vehicle as applied in transport system analysis and planning. Conventionally, total door-to-door travel time enters into the equation for generalised cost and is then converted into monetary units; our model indicates that there should be a distinction between 'normal' travel time from the origin and destination and the extra delay resulting from the flow-dependent nature of travel times. Whether or not cost coefficients for being early or late should also enter into the evaluation criterion for peak-hour transport facilities is a matter for further study and consideration.

There are also implications for traffic assignment practice and highway design, and this too is an important topic of further research. As noted by Teply (1982, p.76): 'Although some past models attempted to include certain elements of feedback, no comprehensive treatment of the problem exists.' Any unmodified application of our model to a network of many origins and destinations would be cumbersome and Alfa (1984) lists possible directions for further work. A procedure for estimating flows and travel times in networks with multiple origin-destination pairs and with time-varying demands is a prerequisite to a generalisation of departure time choice models.

However, the authenticity of our model could be invoked to simplify the assignment process and make it more credible. This could be achieved by taking the total demand and dividing it by the transport bottleneck (or transport corridor) capacity and specifying as the design criterion the temporal duration of the peak. Some subjective assessment of what constitutes a 'reasonable' or 'appropriate' peak period in a given context would be necessary. This suggestion is similar to the normal practice of the structural engineer who divides the total load by the working stress of the structure and executes the design by providing an appropriate cross-section of material. This analogy is highly relevant to the traffic assignment phase of the planning process in much the same way as understanding the mechanisms of peak spread lessens the fear expressed by some people that traffic management techniques will inevitably fail to relieve urban traffic congestion.

A final, long-term application of this work is in the area of optimal control strategies. There are enormous possibilities of remote control of all kinds of activities - including information on the best starting times for a commuter - through the application of electronic devices, but to do this sensibly would involve a better understanding of the behavioural base of the temporal distribution of traffic. The model described in this paper suggests relevant criteria for use in any optimisation program that might make any control strategy less a flight of fantasy.

REFERENCES

Abkowitz, M.D. (1981) 'An analysis of the commuter departure time decision', Transportation, vol. 10, pp. 283-297.

Alfa, A.S. (1979) 'A Model for the Temporal Distribution of Peak Traffic Demand - Development, Calibration and Application to Route Assignment', unpublished PhD thesis, School of Transport and Highways, University of New South Wales.

Alfa, A.S. (1982) 'Parameter estimation for the peak traffic model', Transportation Planning and Technology, vol. 7, pp. 281-287.

Alfa, A.S. (1984) 'Modelling the temporal distribution of peak traffic demand - state of the art and directions', paper presented at Joint ORSA/TIMS Meeting, San Francisco, May.

Alfa, A.S. and D.L. Minh (1979) 'A stochastic model for the temporal distribution of traffic demand - the peak hour problem', Transportation Science, vol. 13, pp. 315-324.

Blunden, W.R. (1982) 'Congestion - friend or foe?', New Zealand Engineering, vol. 37, pp. 13, 15-16.

Blunden, W.R. and J.A. Black (1984) The Land-Use/Transport System (Oxford: Pergamon Press) 2nd ed, Appendix 2, pp. 223-228.

Clarke Gazzard Voorhees (1974) Transportation Planning Eastern District 'B', (Sydney: Council of the City of Sydney).

Clunas, K.J. (1984) 'Transport Disruptions and Travel Behaviour', unpublished MEngSc Project, School of Civil Engineering, University of New South Wales.

De Palma, A., M. Ben-Akiva, C. Lefevbre and N. Litinas (1983) 'Stochastic equilibrium model of peak period traffic congestion', Transportation Science, vol. 17, pp. 430-453.

Fargier, P.H. (1981) 'Influence du mecanisme de choix de l'heure de depart sur la congestion du traffic routier (Effects of the choice of departure time on road traffic conditions)', in V.F. Hurdle, et.al. (eds) Proceedings of the Eighth International Symposium on Transportation Traffic Theory, (Toronto: University of Toronto Press), pp. 223-263.

Gaver, D.P., Jr. (1968) 'Headstart strategies for combating congestion', Transportation Science, vol. 11, pp. 172-181.

Henderson, J.V. (1977) Economic Theory and the Cities, (New York: Academic Press), Chapter 8.

Hendrickson, C. and G. Kocur (1981) 'Schedule delay and departure time decisions in a deterministic model', Transportation Science, vol. 15, pp. 62-71.

Julian, C.R. (1971) 'Staggering work hours to ease existing street capacity problems', paper presented at Institute of Traffic Engineers 1971 World Traffic Engineering Conference, Montreal, September 19-24.

Kuester, J.L. and J.H. Mize (1973) Optimization Techniques with Fortran (New York: McGraw Hill).

Minh, D.L. (1976) 'The Analysis of a Time-inhomogenous Single Server Queueing Model with Compound Poisson Input and General Service Time Distribution', unpublished PhD thesis, School of Transport and Highways, University of New South Wales.

Minh, D.L. (1977) 'Discrete time, single server queue from a finite population', Management Science, vol. 23, pp. 756-767.

O'Malley, B. and C.S. Selinger (1973) 'Staggered work hours in Manhatten', Traffic Engineering and Control, vol. 14, pp. 418-423.

Peat, Marwick, Mitchell and Company (1972) 'An Analysis of Urban Travel by Time of Day', Report, NTIS PB 247-776.

Powell, M.J.D. (1964) 'An efficient method for finding the minimum of a function of several variables without calculating derivatives', The Computer Journal, vol. 7, pp. 155-162.

Remak, R. and S. Rosenbloom (1976) 'Peak-Period Traffic Congestion: Options for Current Programs', National Cooperative Highway Research Program Report, 169 (Washington, D.C.: Transportation Research Board).

Remak, R. and S. Rosenbloom (1979) 'Implementing Packages of Congestion-Reducing Techniques: Strategies for Dealing with Institutional Problems of Cooperative Programs', National Cooperative Highway Research Program Report, 205 (Washington, D.C.: Transportation Research Board).

Rosenbloom, S. (1978) 'Peak-period traffic congestion: a state-of-the-art analysis and evaluation of effective solutions', Transportation, vol. 7, pp. 167-191.

Shallal, L.A.Y. and A.M. Khan (1980) 'Predicting peak-hour traffic', Traffic Quarterly, vol. 34, pp. 75-90.

Small, K. (1982) 'The scheduling of consumer activities: work trips', The American Economic Review, vol. 72, pp. 467-479.

Spearritt, P. (1982) The Sydney Harbour Bridge - A Life (Sydney: George Allen & Unwin).

Teply, S. (1982) 'Network travel behaviour "levels of choice" and urban transportation management', Traffic Engineering and Control, vol. 23, pp. 71-76.

Transportation Research Board (1980) 'Alternative Work Schedules: Impacts on Transportation', National Cooperative Highway Research Program Synthesis of Highway Practice, 73 (Washington, D.C.: Transportation Research Board).

Vickrey, W.S. (1969) 'Congestion theory and transport investment', American Economic Review, vol. 59, pp. 251-261.

Wardrop, J.G. (1952) 'Some theoretical aspects of road traffic research', Proceedings Institution of Civil Engineers (Part II), vol. 1, pp. 325-378.