PERGAMON

Transport Policy 5 (1998) 23–35

TRANSPORT
POLICY

The conceptual structure of traffic jams

Chris Wright*, Penina Roberg[1]

Transport Management Research Centre, Middlesex University, Bounds Green Road, London N11 2NQ, UK

Abstract

Area-wide traffic jams develop through the propagation of queues from link to link, a process that resembles the growth of branches on a tree. The process is not well understood. In this paper, simple models for jam growth arising from a single bottleneck are developed for an idealized grid network. Under these idealized conditions, it has been shown that there are essentially two possible spatial configurations for a traffic jam on the type of network considered, each having a characteristic form sharing some of the properties of a fractal. More important, the models highlight an interesting dilemma in traffic management. A strategy that aims to minimize the rate of growth of a jam by a suitable allocation of queue storage space will actually encourage gridlock at the heart of the congested area, and conversely, a strategy that aims to defer gridlock will result in queues spread over a wider area. Extensive channelization (normally advocated in the interests of efficient traffic flow and safety) will also encourage longer queues. With hindsight, these conclusions seem obvious for any network whether imaginary or real, but they do not seem to have appeared in the literature, and the models give some indication of the size of the effects involved. © 1998 Elsevier Science Ltd. All rights reserved.

Keywords: Traffic congestion; Models; Jams; Dispersal

1. Introduction

In many countries, congestion has become endemic, with traffic jams spreading over large tracts of urban network throughout the working day. Consequently, there is a need to study the process of traffic jam formation and growth in its own right, so that new techniques for controlling traffic jams can be developed. These should include not only the 'high technology' solutions that are currently being researched worldwide, but also various forms of low-cost traffic management measure involving channelization, turn restrictions and traffic signal systems aimed at controlling the development of queues (Huddart and Wright, 1989; Wright and Huddart, 1989a, b).

Here, we propose a simple analytical model for jam growth, compare it with a computer model, and draw some qualitative conclusions about the nature of traffic jams and how they might be inhibited via simple queue management measures.

The networks are idealized, uniform, one-way rectangular grid systems. The theoretical model assumes steady-state demand conditions, whereas the simulation model allows for stochastic variation.

Clearly, our approach sacrifices a great deal of realism, and the results are not intended to be immediately applicable to real networks. However, they yield qualitative insights which may lead to a better understanding of the congestion problem in general terms. In particular, they predict certain features of jam behaviour that we did not expect, and point to low-cost strategies for traffic management.

1.1. Sources of traffic jams

A traffic jam can start in one of three ways:

1. A temporary obstruction.
2. A permanent capacity bottleneck in the network itself.
3. Stochastic fluctuation in demand within a particular sector of the network, leading to spillback and queue propagation.

We are concerned only with jams arising from a single source that might be categorized under (1) or (2). Whichever of these applies, one would expect the subsequent process of jam development to be the same.

1.2. Spillback

The discharges of vehicles into the various exits from a road link are unlikely to be independent, because (a) vehicles making different turning movements may share the same traffic lanes, and (b) even if there is a separate

* Corresponding author.

[1] Now Penina Orenstein, Bellcore, Redbank, New Jersey, USA.

0967-070X/98/$19.00
PII: S0967-070X(98)00006-7

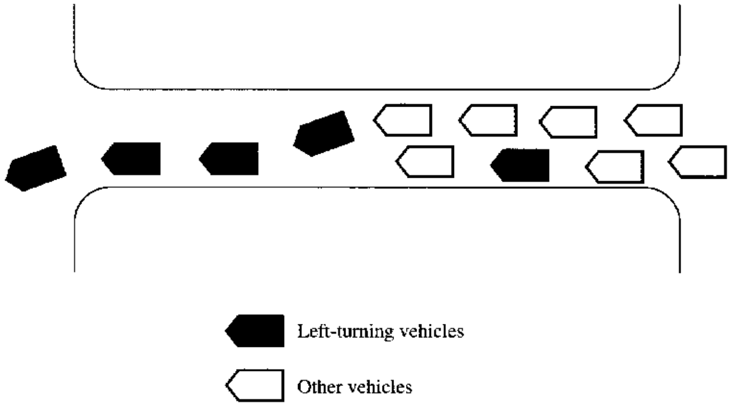

Fig. 1. Interference between turning vehicles and ahead vehicles on a road link where the turning discharge is obstructed.

turning lane or lanes for each exit, drivers do not necessarily position themselves in the correct lane at the entrance to the link under consideration—there is usually a certain amount of weaving between the entrance and the exit of a link during which the different turning movements interfere with one another's progress.

If a particular exit (say, the left-turn exit) is blocked, vehicles intending to turn left will form a queue that spills back along the link. If lane discipline is not perfect, the queue may eventually spread across the other lanes and block all the traffic (Fig. 1).

Very little is known about the mechanism of interaction between queues in practice. Here, we shall picture a road link as divided into two distinct zones: a downstream queue storage area where vehicles are organized into separate turning movements, and an upstream 'reservoir' where the turning movements are mixed. The interactions occur at the transition between these two zones.

1.3. Queue propagation

Starting from an initial obstruction located somewhere near the centre of the study area, we visualize a traffic jam as propagating from link to link via a branching process, each queue generating new branches at each junction in turn. Initially, the topology of the branches resembles that of a simple tree, but at some stage, a queue will tail back around four sides of a block to form a closed loop. This may be repeated elsewhere, with additional closed circles, so that the tree evolves into a form that would more accurately be described as a 'lattice'.

Note that the pattern of queue propagation in our model depends only on the interference between turning movements, not on the origin–destination movements.

1.4. Quantifying the severity of a traffic jam

It is possible to quantify aggregate journey times and delays within our models, but we have found it easier to use simpler measures. The first is the time taken for gridlock to develop following an obstruction (the longer the time taken, the better). The second measure is the rate of growth of the traffic jam as a whole, where the size of the jam at any particular moment is represented in terms of the number of city blocks enclosed within the jam boundary. By 'jam boundary', we mean an imaginary polygon joining the outer extremities of the queues at any particular moment in time. This is not a rigorous concept, but it is intuitively a useful one when dealing with jams that are spread out over a large area of network.

2. An analytical model for queue propagation on one-way rectangular grid networks

2.1. Model assumptions

The model networks consist of two sets of parallel one-way roads at right angles, with the direction of traffic alternating between neighbouring roads, and the roads at equal spacing. There are no intermediate origins or destinations within any of the links, and only the severed links on the boundary act as traffic sources and sinks.

Vehicles travelling along any given link may choose from two possible movements at the downstream junction: 'ahead' and 'turning'. The turning movements alternate between left and right turns at successive junctions along any given road.

Conversely, at any given junction, vehicles may arrive via either of two approaches. These have equaly priority; by this we mean that the opportunities for discharge are shared between the approaches in proportion to the demands. Now, in our analytical model, the discharge rate across each stopline is constant: it changes only when a queue spills back over the stopline from downstream, reducing the flow instantaneously to a lower but constant level (which may be zero). In effect, therefore, we are picturing each junction as controlled by a traffic signal

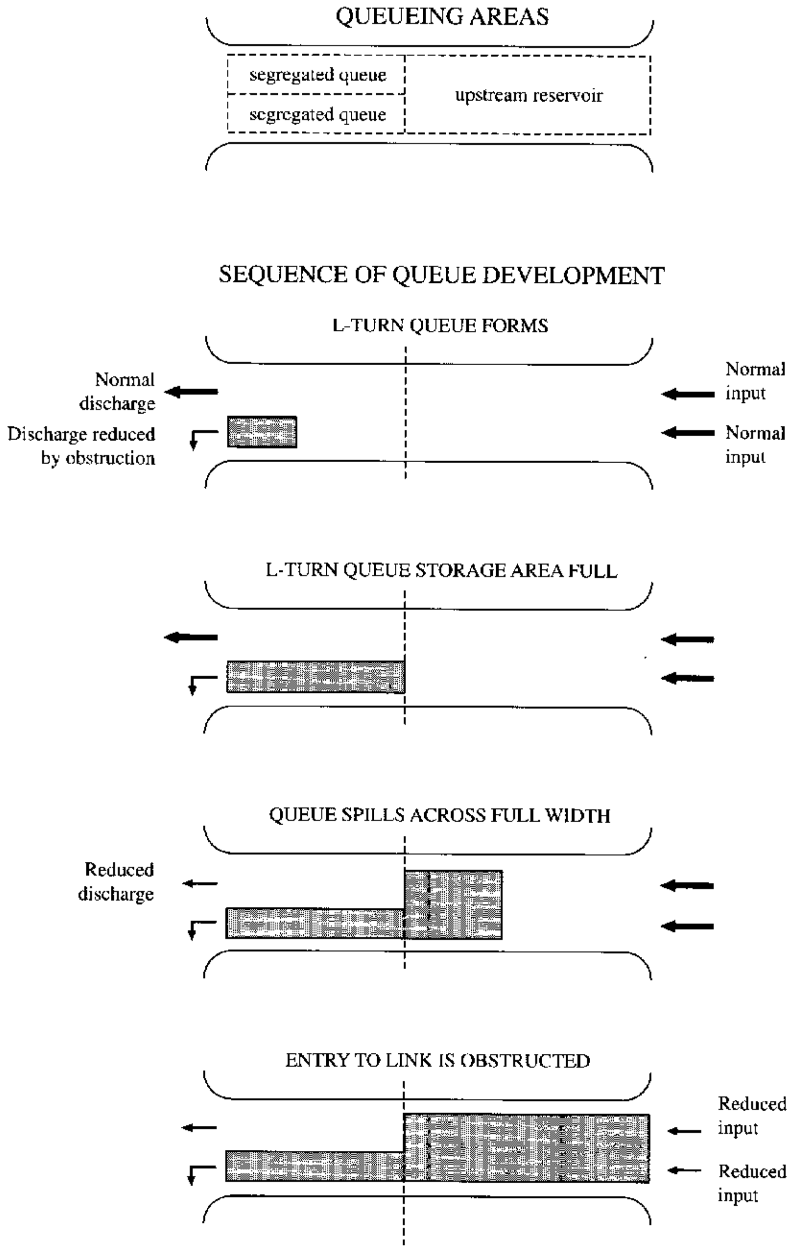

Fig. 2. Assumed sequence of queue spillback within a one-way road link.

which has very short cycle times, and is undersaturated for all approaches.

We assume that there is no 'cross-blocking', i.e. the spillback does not block crossing traffic at the upstream junction. Instead, vehicles wait on their respective approaches until there is space for them to proceed. (In practice, drivers are not always so well disciplined, but we have not pursued this aspect here.)

Each link contains just two segregated queue storage areas terminating at the downstream stopline: one for the 'ahead' movement and one for a turning movement (see Fig. 2). Each is of constant width; it may consist of one or more lanes, but there are no shared lanes. Upstream, there is a 'reservoir' that feeds into both of them. It occupies the full width of the road. One can visualize the reservoir as being channelized (i.e. it may possess lane markings), but within this area, vehicles are allocated randomly between the lanes regardless of their intended direction of movement at the downstream junction. The mechanism of queue spillback within these storage areas will be described later.

When a link fills up with vehicles, the jam will propagate to further links immediately upstream. Additionally, the following assumptions are made:

- The flow between any origin–destination pair is constant. Furthermore, the route used by each driver is effectively fixed in advance, and in particular, drivers do not change their routes in response to congestion.
- For the purpose of the analytical model, only the deterministic component of flow is taken into account, i.e. we ignore stochastic variations, together with any cyclic variations caused by alternate red and green periods at traffic signals.
- In the absence of any obstruction, the 'ahead' flows at all the stoplines are the same, and the turning flows at all the stoplines are the same irrespective of whether they are turning left or right (but not necessarily equal to the ahead flows).
- When unobstructed, the stopline capacities of the various junction approaches, and individual lanes within the junction approaches, are fixed, and proportional to the width of road occupied by the lanes under consideration. Specifically, the capacity per unit width is the same for all 'ahead' queues, and the capacity per unit width is the same for all turning queues (but not necessarily the same as for the ahead queues).
- The width of the upstream reservoir on each link is uniform along its length and the same for all links. Similarly, the width of the ahead queue storage area is uniform and the same for all links, and the width of the turning queue storage area is uniform and the same for all links. The total stopline width (i.e. the ahead and turning storage areas combined) is greater than or equal to the upstream reservoir width.
- The density of vehicles per unit area of queue storage space is the same for all queues.
- The speeds of all vehicles are assumed to be the same everywhere on the network until they enter spillback queues. When a vehicle enters a queue, its speed instantaneously falls to that of the other vehicles in the queue, which is controlled by the discharge rate of vehicles at the head of the queue.

2.2. Notation

We shall use the following notation:

t_A	Time taken for a capacity restriction affecting the ahead discharge from a link to propagate to the upstream end of the link
t_{LR}	Corresponding time taken for a capacity restriction affecting the turning discharge from a link to propagate to the upstream end of the link
q_A	The demand for ahead movement at the downstream exit from each link when there is no obstruction to flow, in vehicles/unit time
q_{LR}	The demand for turning movement (i.e. left or right turns) at the downstream exit from each link when there is no obstruction to flow, in vehicles/unit time
q	Total demand on each link (equal to $q_A + q_{LR}$)
c	Ratio of discharge capacity to demand for a link
N	Total number of vehicles that can be stored in a link
α	Proportion of a segregated queueing area devoted to *ahead* queue storage (the proportion devoted to storage of the turning queue will then be $1-\alpha$)
ϕ	Equal to q_A/q: proportion of vehicles travelling in the ahead direction
σ	Proportion of the area of a link devoted to storage of segregated queues (the proportion devoted to the upstream reservoir will then be $1-\sigma$). Note that $0 < \sigma < 1$.

2.3. Link propagation times

An important characteristic of the system is the time taken for a queue to propagate along the whole of a link. If this link propagation time is small, the jam will escalate quickly.

Now, a downstream obstruction may impinge on the link in either of two distinct ways:

1. The *ahead* discharge flow is reduced from q_A to cq_A, when $c < 1$.
2. The *turning* discharge flow is reduced from q_{LR} to cq_{LR}, when $c < 1$.

If the ahead discharge flow is reduced, a queue will build up until it fills the ahead storage area. It is then assumed to spread across the whole width of the road, blocking the upstream reservoir. Within this area, ahead and turning vehicles will be mixed together in proportion to their respective demands. If follows that the throughput of all vehicles will now be reduced by the factor c. The queue in the reservoir will in turn build up and spill back to the upstream entry. This sequence of queue spillback and blocking is shown diagramatically in Fig. 2. The overall link propagation time, t_A, will be the sum of the individual spillback times for the ahead storage area and the upstream reservoir.

On the other hand, if the turning discharge is reduced, the turning storage area will fill up first, and then the upstream reservoir, in which case the overall link propagation time t_{LR} will be the sum of individual spillback times for the turning storage area and the upstream reservoir.

We can now derive expressions for t_A and t_{LR} in terms of the stopline width allocation parameter α and the segregated proportion parameter σ. We can write, for example,

$$t_A - t_{LR} = \frac{N}{q(1-c)} \times \frac{\sigma(\alpha - \phi)}{\phi(1-\phi)} \tag{1}$$

See Appendix A for the derivation of this result.

2.4. Spatial structure of the idealized traffic jam

Once a link has filled up with queuing traffic, its capacity to accept vehicles from either of the two upstream links feeding into it will be affected in the same way: it will also be reduced to a value c times the demand. Effectively, the initial obstruction is propagated to the two upstream links, which will now behave in a similar way.

The queues branch out along the quickest propagation 'routes' to form a queue propagation 'tree', which here takes on a regular geometrical form. To find its shape, assume that the initial obstruction occurs at the downstream end of a particular link, close to the stopline. It may affect

1. The 'ahead' vehicles,
2. The turning vehicles, or
3. Both.

Note that, after the first link is full, the subsequent pattern of queue propagation will be the same for all three of the above cases: the only element which varies is the time taken for the first link to overflow.

However, the growth paths are affected by the relative values of the link propagation times t_A and t_{LR}. It is possible to distinguish between three possible configurations, corresponding to the following three conditions:

Type I: $t_A < t_{LR}$
Type II: $t_A > t_{LR}$
Type III: $t_A = t_{LR}$

From Eq. (1), it is clear that these conditions are respectively equivalent to:

$$\alpha < \phi \quad (2)$$

$$\alpha > \phi \quad (3)$$

$$\alpha = \phi \quad (4)$$

Condition (4) represents the case where the segregated queue storage area is allocated between the ahead and turning streams in eactly the same ratio as the demands: this might be referred to as a 'balanced' layout. Mathematically, it represents a degenerate case.

Conditions (2) and (3), on the other hand, represent 'unbalanced' layouts. Condition (2) implies that proportionately less storage is devoted to the ahead traffic; given that its discharge is restricted by a given factor c, a blocked ahead queue tails back more quickly than a blocked turning queue. Condition (3) implies that proportionately less storage is devoted to the turning traffic, so that a blocked turning queue tails back more quickly than a blocked ahead queue.

It follows that, in the case of Type I jams, queue growth takes place most quickly in a straight line along one axis of the grid network. Other, secondary queues branch out in straight lines at right angles from this primary queue, and yet more queues branch out from the secondary queues, and so on. The resulting propagation paths are shown in Fig. 3. The pattern is dominated by four primary paths (shown as heavier lines) that radiate from the source in four compass directions. They divide the system into four quadrants, each having a distinct 'grain' to its growth pattern.

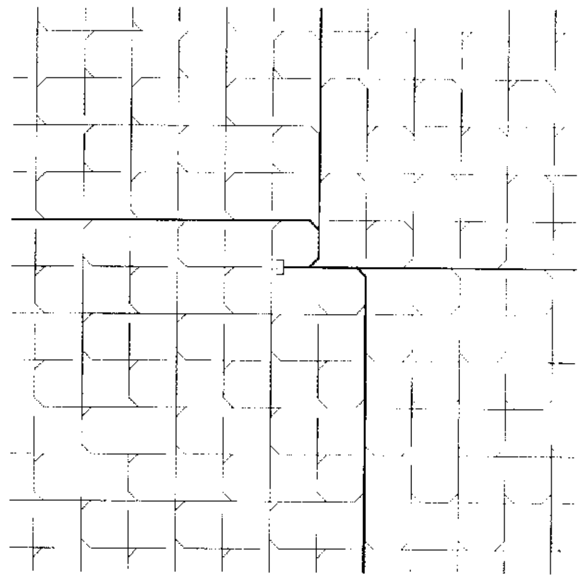

Fig. 3. Queue propagation paths for the case where $t_A < t_{LR}$ (Type I jams).

In the case of Type II jams, the queues form small clusters of tightly packed 'curls', and as shown in Fig. 4, the overall pattern can be divided this time into eight sectors, each having a distinct 'grain'.

An example of a Type I traffic jam is illustrated in Fig. 5, which shows the progress of queues over a period of 10 time units after the occurrence of the original blockage, for a grid

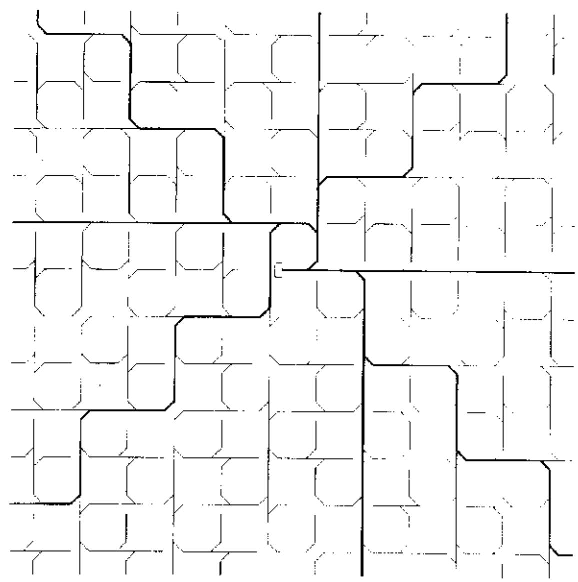

Fig. 4. Queue propagation paths for the case where $t_A > t_{LR}$ (Type II jams).

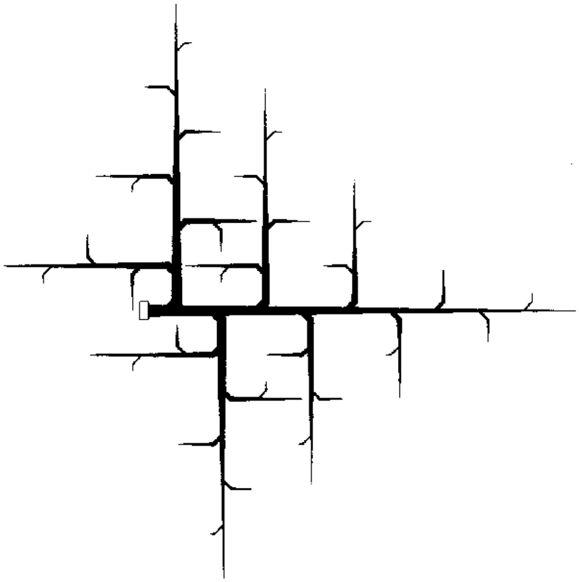

Fig. 5. Congestion lattice with $t_A = 1$ unit and $t_{LR} = 3$ units (Type I jam). A total of 10 time units have elapsed since the original obstruction.

where the ahead link propagation time is 1 unit and the turning propagation time 3 units. An example of a Type II jam is illustrated in Fig. 6, for a grid where the ahead link propagation time is 3 units and the turning propagation time 1 unit. In each case, the thickness of the line represents the time elapsed since the occurrence of the initial blockage.

In tracing out these growth paths, we have ignored the possibility that traffic feeding into an obstructed link may itself be reduced below the normal level if vehicles are held up in another part of the jam. We refer to this as traffic 'starvation'. Later, using a computer simulation model, we shall see that starvation can change the shape of the jam boundary considerably.

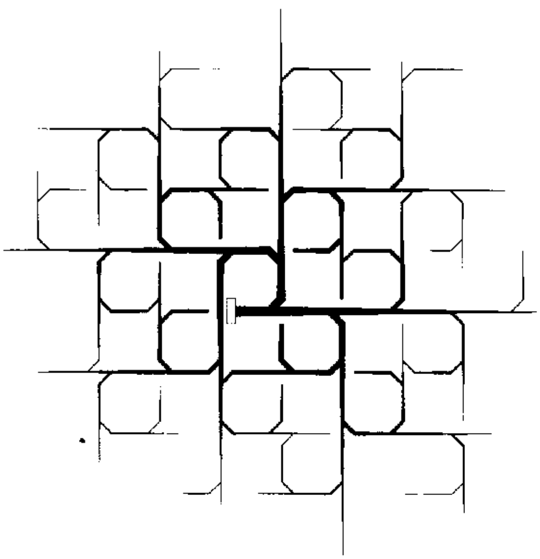

Fig. 6. Congestion lattice with $t_A = 3$ units and $t_{LR} = 1$ unit (Type II jam). A total of 16 time units have elapsed since the original obstruction.

3. Controlling jams

We now assess the effects of changing some of the network parameters on jam growth. The first step is to develop simplified expressions for jam growth rates for the asymptotic case, where the diameter of the jam is large in comparison with the city-block size. This is the topic of the next section. Subsequently, we consider the effects of different queueing space configurations on (a) the rate of expansion of the jam boundary and (b) the propensity to develop gridlock.

It is not difficult to show that, when considered over a large area of network, the boundary of our idealized traffic jam has asymptotically a polygonal shape. We refer first to Fig. 3, which shows the paths along which queues propagate for a Type I jam. For any point in the network, one can find a quickest growth path from the initial obstruction to that point which involves, at most, four turns. For points distant from the initial obstruction, the time required for the queue to negotiate these turns becomes a negligibly small proportion of the total, and the overall time required for a queue to reach that point will be asymptotically equal to the sum of the E–W and N–S distances of the point from the obstruction, measured in blocks, multiplied by the 'ahead' link propagation time.

It follows that the boundary has a diamond shape centred on the initial obstruction, with the two diagonals asymptotically equal in length and parallel to the two sets of road links, as shown in the upper diagram of Fig. 7.

For Type II jams, we refer to the diagram shown in Fig. 4. For each point on the network, one can find a quickest queue propagation path consisting of an equal number of consecutive left and right turning movements coupled with a straight line segment, together with two additional turns, at most. Asymptotically, the boundary is octagonal, as shown in the middle diagram of Fig. 7. It is not a regular octagon, although it has four axes of symmetry. The octagon approaches the shape of a square as the ratio α/ϕ approaches infinity, as shown in the lower diagram of Fig. 7.

The Type III jam, a degenerate case, represents the borderline between Types I and II.

3.1. Minimizing jam growth rate

We can now vary the network parameters to see the effect on the rate of traffic jam growth. The parameters include the proportion of link area that is devoted to segregated storage space (σ), and the allocation of stopline widths between ahead and turning traffic (α).

We shall consider Type I jams first. The initial queue propagates upstream from link to link in the opposite

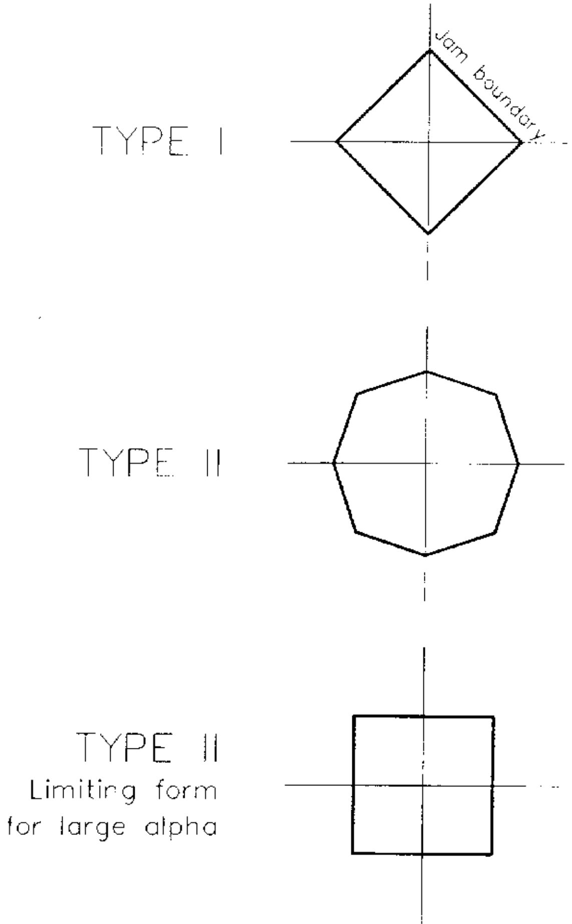

Fig. 7. Alternative configurations of jam boundaries after a long period.

direction to the ahead flow, and the tail of this queue defines one corner of the boundary. After an elapsed time τ, the number of links covered by this queue will be τ/t_A.

The diagonal of the square will be twice this distance, and the area (denoted by A) of the traffic jam at time τ will be given by

$$A = 2\tau^2/t_A^2 \tag{5}$$

blocks. Clearly, in order to minimize the rate of growth of the traffic jam, we need to maximize the value of the ahead link propagation time.

In Appendix A, we show that this link propagation time is given by the expression

$$t_A = \frac{N}{q(1-c)}\left[\frac{\sigma(\alpha-\phi)}{\phi} + 1 - \frac{qT}{N}\right]$$

which, when substituted into Eq. (5), gives

$$A = \frac{q^2(1-c)^2}{(N-qT)^2}J\tau^2$$

where

$$J = 2/(1+\Gamma)^2 \text{ with } \Gamma = \frac{N\sigma(\alpha-\phi)}{(N-qT)\phi}$$

and $0 < \alpha \leq \phi$.

The quantity J is a constant, which we refer to as the 'jam area coefficient'. The larger the coefficient, the faster the jam grows.

Now, under appropriate conditions, α could lie anywhere within the closed interval (0,1). By considering all the possible combinations of α values that could occur together with the other parameters, one can show that the lowest possible value that Γ could assume is -1.

The smallest possible value of J for Type I jams, and hence the smallest rate of jam growth, occurs precisely when $\Gamma = 0$. This is because $\alpha < \phi$ and Γ never assumes a value less than -1. This condition will be met if either

(a) $\alpha \to \phi$, i.e. a 'balanced' layout in which the widths allocated to the segregated approach queues are in exactly the same ratio as the demands, or
(b) $\sigma = 0$, i.e. the area and hence the length of road link over which the queues are segregated is equal to zero.

A similar analysis can be carrier out for Type II jams. It yields a more complicated expression for the jam area coefficient, but over the range of conditions likely to be met in practice, the jam growth rate is again minimized (or nearly so), when $\Gamma = 0$. The two configurations (a) and (b) are alternative ways of achieving the same thing: they both avoid wasted space in the 'ahead' and 'turning' areas that cannot be filled because of queue interference between arriving vehicles. Consequently, they delay spillback to the upstream junction.

This conclusion applies to any road link in any network, not just the idealized models described here.

So, it seems to be desirable to aim in practice for

- a balanced layout insofar as it can be achieved, together with
- the minimum length of segregated queue storage consistent with the need for lane discipline under normal operating conditions.

3.2. Inhibiting gridlock

We now turn to the question of gridlock, where the situation is quite different. Consider a set of four links that surround a city block, and suppose that an obstruction occurs at the stopline on one of these four links. A queue will form and eventually propagate to the other three links in turn, so that the tail of the queue eventually arrives back at the starting point of the jam.

Once this has occurred, the jam may be much more difficult to clear than would otherwise be the case. Let us refer to the time required as the 'gridlock time'. Initially, the obstruction may affect only the turning traffic on the link under consideration, or the ahead traffic, or both simultaneously. We will assume that it affects the turning traffic: the other cases differ only in detail and the overall conclusions are the same. The gridlock time, denoted by t_g, is given by

$$t_g = 4t_{LR} \tag{6}$$

However, from Eq. (0) we have

$$t_{LR} = \frac{N}{q(1-c)}\left[\frac{\sigma(\phi-\alpha)}{(1-\phi)}+1-\frac{qT}{N}\right]$$

which, when substituted into Eq. (6), gives:

$$t_g = \frac{4(N-qT)}{q(1-c)}G$$

where G, which we refer to as the 'gridlock time coefficient', is a function of the layout parameters, and is given by

$$G = 1 - \frac{\phi\Gamma}{(1-\phi)}$$

Other things being equal, we would wish to postpone gridlock for as long as possible, and hence maximize G. Now, by definition the proportion of ahead vehicles ϕ satisfies the condition $0 < \phi < 1$, so that $\phi/1 - \phi > 0$. It therefore follows that G is the greatest when Γ is large and negative. This implies that $\alpha < \phi$. In other words, proportionately more stopline width is allocated to the turning traffic, so that $\alpha - \phi$ is strictly less than zero. We must force a Type I traffic jam, and it will be advantageous to set $\sigma = 1$; in other words, to segregate the ahead and turning queues along the whole length of each link. Qualitatively, it is easy to see that to postpone gridlock we reserve as much storage space as possible for turning vehicles, thus reducing the rate at which spillback proceeds around the four sides of the block.

Hence, there is a direct conflict between the requirement to maximize the gridlock time and the requirement to minimize the overall growth rate of the jam.

4. The computer simulation model

The above model does not take into account randomness in the vehicle arrival times, nor the 'starvation' phenomenon mentioned in Section 2.4. To see how these factors might affect jam structure, we turn to a simulation model. It is based on the same assumptions about queueing behaviour as the theoretical model, with traffic fed into the roads at the edge of the grid (details are given in Appendix A.3).

4.1. Results from the simulation model

Three main series of tests were carried out using the simulation model. The first was concerned with the rate of jam growth over time, to test the hypothesis predicted by the analytical model that it would be desirable to aim for a 'balanced' layout, where the widths of the segregated approach queues are allocated in proportion to the demands ($\alpha = \phi$), with minimal length of segregated queue storage (small σ). The second was concerned with delaying the onset of gridlock.

4.1.1. Minimizing jam growth rate

A number of simulation trials were conducted, and for each run, the size of the traffic jam (measured in terms of the total number of blocked links) was recorded after 11 cycles. N was set at 50, the demand was 24 (vehicles/min), the capacity was 50 (vehicles/min) and the proportion of vehicles travelling in the ahead direction (ϕ) was 0.8. The parameters that varied from trial to trial were the stopline width allocation parameter α and σ, the parameter which monitored the proportion of link area devoted to the segregated queues. Thus, α and σ satisfied the conditions $\alpha \in \{0.33, 0.5, 0.66, 0.75, 0.80, 0.85\}$ and $\sigma \in \{0.4, 0.5, 0.6, 0.7, 0.8\}$.

Fig. 8 shows the effect of segregated queue storage capcity on the jam size for various channelization regimes. The results indicate that the rate of jam growth decreases proportionately with σ. In addition, the rate of jam growth is minimized when α approaches 0.8. Thus, when $\alpha = \phi = 0.8$, the jam expansion rate is minimized. These results confirm the fact that the jam expansion rate can be curbed if the stopline widths are allocated in proportion to the demands, with minimal segregation.

4.1.2. Inhibiting gridlock

The number of cycles until the onset of gridlock was recorded for various values of the stopline allocation parameter, α. The model parameters were the same as in the previous experiment, except that the proportion of area devoted to the segregated queues was fixed at 0.4 throughout.

The results confirm the analytical model prediction that the onset of gridlock can be delayed by reducing α, i.e. by allocating a larger proportion of stopline width to the turning traffic. Fig. 9 shows the effect for the particular parameter values listed above; however, a similar pattern was observed for other demand levels as well.

4.2. Queue starvation

Fig. 10 shows that, with the simulation model, the boundary is roughly diamond shaped, as expected. However, some of the links inside the boundary become 'starved' of vehicles, and do not accumulate queues at all. Since the density of vehicles on these links is less than normal traffic, we refer to them as 'anti-queues'.

Starvation may affect the jam boundary as well, but this only becomes apparent for large jams. Fig. 11 shows a simulated traffic jam about the size of London. The picture was achieved by running the simulation model over a 256 × 256 one-way grid network with $\alpha = 0.5$ and $\phi = 0.8$, parameters for a Type I jam. Each dot represents a blocked

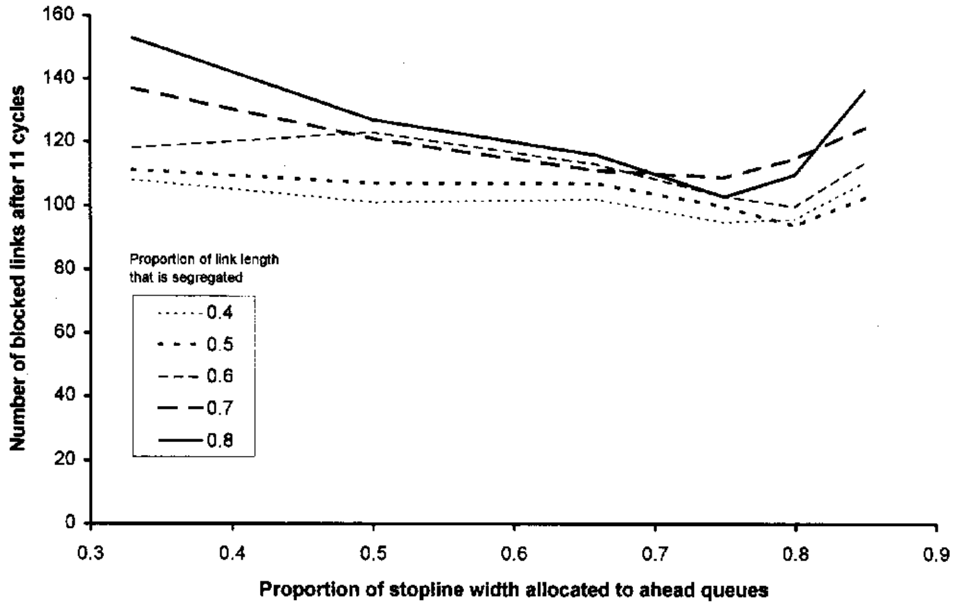

Fig. 8. Effect of segregated queue storage on jam size for various channelization regimes.

junction. The anti-queues are not shown. The jam boundary still exhibits the overall diamond shape, but it is indented like a four-petalled flower. In fact, this is a fractal and its structure has been analysed in detail in an earlier paper by (Abbess and Roberg, 1995).

Starvation has a third effect that is less immediately obvious than the two already mentioned: because vehicles are not able to leave the area enclosed by the jam boundary at the rate they would otherwise have done, outbound flows are depleted for some distance outside. The 'halo' of reduced traffic flows around the boundary is just perceptible in some of the computer visualizations.

The jam process can also be viewed as one of diffusion-limited aggregation (DLA; see Abbess and Roberg, 1995), because the rate of jam growth at the centre is faster than the rate at the tips: in other words, the growth pattern is dendritic. [The reader is referred to (Witten and Sander, 1981), for a general treatment of DLA processes and (Batty, 1990) for a description of their application to the development of urban structure.]

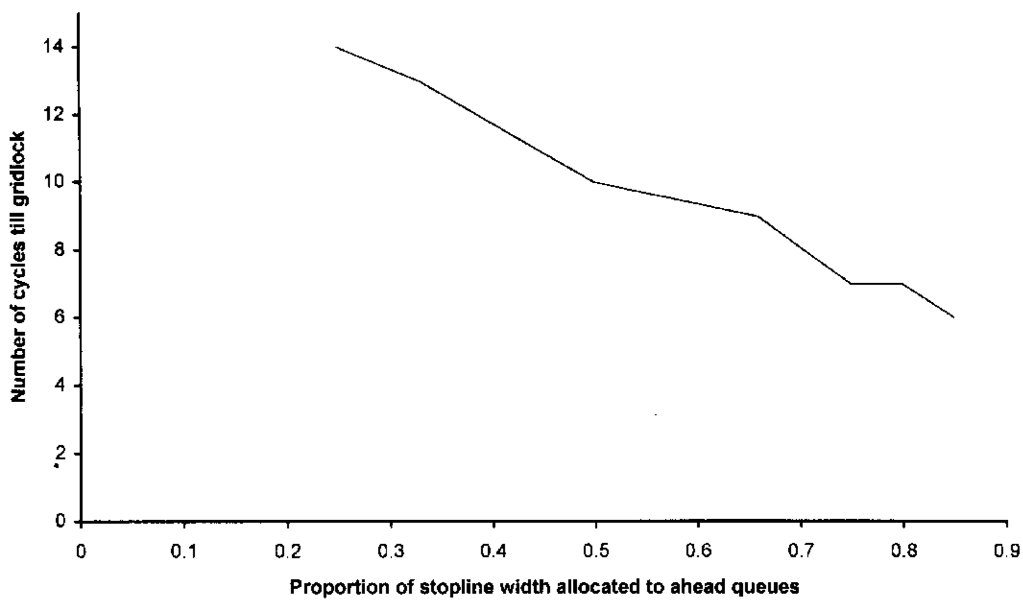

Fig. 9. Effect of channelization on time till gridlock.

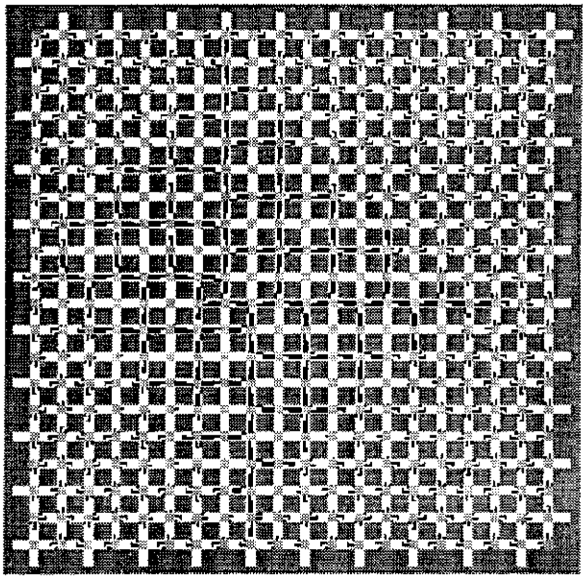

Fig. 10. Spatial characteristics of a simulated Type I jam on a 20 × 20 one-way network.

4.3. Effects of stochastic variation in vehicle arrival rates

At the local level, random fluctuations in traffic flow can have a marked effect on queue lengths, a fact which the analytical model does not take into account. Consequently, at an early stage the computer simulation model was adapted to run with Poisson flows feeding into the edge of the grid. This seemed to have little effect on the overall shape and size of the traffic jam after the first few cycles, except that the boundary was slightly less regular in shape.

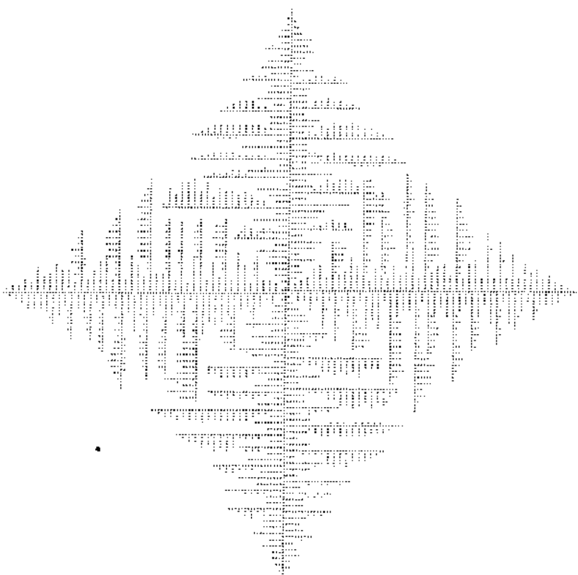

Fig. 11. A large simulated traffic jam on a 256 × 256 one-way grid network.

Unlike traffic signal queues under light demand conditions, traffic jams occur as an accumulation of relatively large numbers of vehicles over a relatively long period of time, and as the size of the jam increases the mean rate of accumulation increases to a level such that random fluctuations from one time period to the next are small compared to the mean, even with Poisson input flows. The results are not sensitive to the precise way in which initial queues form close to the obstruction.

5. Conclusions

A simple analytical model, supplemented by a simulation model, has been used to demonstrate how the rate of growth of a jam, and its propensity to develop gridlock, vary with the queue storage arrangements. The model ignores the effects of route choice, since the overall pattern of jam development is governed not so much by the routes of vehicles but by the overall turning proportions at each junction in the network. Whilst this approach may sacrifice some realism, some general principles have emerged that may be applicable to more realistic conditions.

The key results may be summarized as follows. When an obstruction is created at a particular point on a grid network of the type described, a queue forms and subsequently branches out over a wide area. The theoretical analysis predicts that the boundary will have either a diamond shape or octagonal shape in plan. As one might expect, the rate of jam growth is affected among other things by the severity of the original blockage and the overall level of traffic demand. It is also affected by the proportion of link storage area denoted to the segregated queues and the balance of storage area between ahead and turning queues.

In addition, the allocation of storage areas affects the time required for gridlock to occur. However, the onset of gridlock and the growth rate of the traffic jam as a whole are to some extent independent. They may be influenced in different ways by the various network parameters: strategies that yield improvements in one respect may not yield a similar improvement in the other. In particular, there are circumstances in which channelization postpones gridlock but can actually increase the rate of growth of the jam and hence the overall delay.

What is less obvious is that, in addition, vehicle starvation can alter the jam structure in three ways: firstly by producing anti-queues, secondly by creating wider variations in the shape of the jam boundary and thirdly by creating a halo of reduced traffic flow outside the congested area. The anti-queues are particularly prominent in Type 1 jams, and are always oriented away from the jam centre. They may have a practical use in clearing jams once they have formed, via a routeing strategy in which vehicles are diverted at right angles to their original paths until they clear the jam boundary, and later return to their intended paths.

The authors have since extended the simulation model to deal with a two-way grid network and, in more recent work, have reproduced the model for small networks using CONTRAM and SATURN. Preliminary results indicate a similar pattern of jam development despite differences in the modelling assumptions. Ultimately, we envisage that the model could be used as a tool for diagnosing and prescribing treatment for trouble spots in central urban areas (Roberg and Abbess, 1995).

Further work is needed to resolve some outstanding issues. First, empirical studies of queue interaction and cross-blocking would be desirable. Second, these phenomena need to be represented in the model in a more realistic form, and tests made to establish the sensitivity of the model results to the assumptions made. However, real jams are surprisingly difficult to observe, and the real environment is so complex that local events can mask what is really going on. All we can say for the moment is that theoretical considerations point to cross-blocking as not being very important once a jam has started to form, although it can markedly affect the gridlock time. Local variations in the interaction mechanism between ahead and turning queues on each link are probably more important.

We suspect that, qualitatively at least, the practical implications of our model results are relevant over quite a wide range of real life conditions, including traffic jams caused by permanent bottlenecks in the road network. They are very simple and (with hindsight) fairly obvious. First, a traffic jam will grow more quickly if vehicles cannot fill up the potential queue storage space on each link. Accurately matching the allocation of turning queue lane width and ahead queue lane width to the corresponding demands will reduce the likelihood of empty spaces and keep the jam compact. (Other measures such as the prevention of cross-blocking are also desirable.) Of course, this is difficult if the proportion of turning traffic varies over time, but the effect of any imbalance can, in theory, be minimized by making the channelized lanes as short as possible.

On the other hand, if the first priority is to postpone gridlock, lane width should be allocated preferentially to the turning queues, so that they back up less speedily around the sides of a block.

It seems that one cannot do both, and it is interesting to consider which of the two strategies might work best. This problem, together with the broader questions of whether congestion management measures actually help or hinder the delivery of a sustainable transport system in the long run, will be considered in a separate paper.

Acknowledgements

The authors gratefully acknowledge the support of the Rees Jeffreys Road Fund and the UK Engineering and Physical Sciences Research Council for this project.

Appendix A

Appendix A.1 Queue spillback times

Any road link in the network is divided into three distinct storage areas. It is useful to derive a basic result that applies to each of them. Suppose that, owing to congestion downstream, the discharge rate from the storage area in question is reduced to a value equal to a constant c times the demand, where $c < 1$, causing a queue to form. Eventually, the queue will grow until it fills the storage area.

Suppose the storage area is of length L', the maximum number of vehicles that can be stored in it being N'. Let the journey time of vehicles through the storage area when vehicles are moving normally under unobstructed conditions be T', and let the demand (i.e. the input flow at the upstream end) be q' vehicles per unit time.

We define the time origin as the moment when discharge is initially obstructed. There will be $q'T'$ vehicles in the area at this moment. Let the time taken for a queue to develop and fill the storage area be t'. Then, the number of vehicles in the queue at the moment when the storage area fills will be equal to the number in the area at time zero plus the number entering during the interval t', minus the number leaving during that interval. It follows that

$$N' = q'T' + q't' - cq't'$$

Rearranging, we get

$$t' = \frac{(N' - q'T')}{q'(1-c)} \tag{A1}$$

The value of t' represents the time taken for a capacity restriction to propagate from one end of a queue storage area to the other.

Appendix A.2 Link propagation times: derivation

Referring to Fig. 2, let N_A and N_{LR} respectively be the maximum number of vehicles that can be stored in the *ahead* and *turning* segregated areas. Also, let N_F denote the maximum number of vehicles that can be stored in the reservoir portion of the link. Then N, the maximum number of vehicles that can be stored on a link, is equal to $N_A + N_{LR} + N_F$.

Now, let T_A be the journey time of vehicles through an *ahead* storage area, when there is no obstruction to the flow, and let T_{LR} and T_F be the corresponding journey times through the turning segregated storage area and the upstream reservoir respectively. Since the ahead storage area has the same length as the turning storage area, we assume T_A and T_{LR} are equal. Also, T, the journey time of vehicles along a whole link when there is no obstruction to flow, is equal to $T_A T_F$ or $T_{LR} + T_F$.

Using Eq. (A1), we can now evaluate each of the two propagation times t_A and t_{LR} as the sum of the relevant segregated storage area spillback time and the reservoir spillback time thus:

$$t_A = \frac{(N_A - q_A T_A)}{q_A(1-c)} + \frac{(N_F - qT_F)}{q(1-c)}$$

and

$$t_{LR} = \frac{(N_{LR} - q_{LR} T_{LR})}{q_{LR}(1-c)} + \frac{(N_F - qT_F)}{q(1-c)}$$

Now, since $N_A = \alpha\sigma N, N_{LR} = (1-\alpha)\sigma N$, $N_F = (1-\sigma)N$, $q_A = \phi q$, $q_{LR} = (1-\phi)q$ and $T_A + T_F = T_{LR} + T_F = T$, it follows that

$$t_A = \frac{N}{q(1-c)}\left[\frac{\sigma(\alpha-\phi)}{\phi} + 1 - \frac{qT}{N}\right] \quad (A2)$$

and

$$t_{LR} = \frac{N}{q(1-c)}\left[\frac{\sigma(\phi-\alpha)}{(1-\phi)} + 1 - \frac{qT}{N}\right] \quad (A3)$$

A little algebra yields

$$t_A - t_{LR} = \frac{N}{q(1-c)} \times \frac{\sigma(\alpha-\phi)}{\phi(1-\phi)}$$

Note that the above equation relies on the assumption that the journey time through the turning segregated storage area and the corresponding time through the ahead storage area are the same. However, even if they were different, the effect would only be to add a constant term to the equation. This does not alter the behaviour of the model fundamentally.

Appendix A.3 The simulation model

The simulation model differs only in detail from the analytical model. First, the flows are not continuous: time is divided into a series of equal slices, and for each of the three queue storage areas within each link (i.e. the upstream reservoir, the ahead queue storage area and the turning queueing storage area), a record is kept of the total inputs and outputs during each time slice. At the end of each time slice, the inputs and outputs are compared and the queue sizes are updated accordingly. A typical model cycle involves all of the east–west roads being processed in turn, starting with the link at the downstream end and moving progressively upstream from link to link. Then, all the north–south roads are processed. Initially, this cycle is carried out several times without any obstruction being placed on the network, to allow the system to stabilize. An obstruction is then introduced on a link near the centre of the grid, a jam evolves from the obstruction and the cycle is repeated until the jam reaches the edge of the grid.

Second, during each time slice, the outputs from each storage area are taken as the inputs to the appropriate storage areas immediately downstream for the next time slice. The simulation model therefore takes account of the starvation phenomenon.

Third, the model addresses the movements of individual vehicles. However, the routes that they follow are not predetermined, and it does not 'track' them through the system. Suppose that X vehicles are discharged from any particular reservoir during a given time slice. They are split into two groups: the number of vehicles allocated to the 'ahead' group is the nearest whole number to ϕX, while the remainder are allocated to the turning group. The two groups are progressed respectively to the downstream ahead and turning queues, regardless of origin. Effectively therefore, individual vehicles perform a random walk over the network, with the probability of moving ahead at any junction equal to ϕ. This represents a simplification in conventional traffic modelling similar to that adapted by Holden and Risebro (1995) and Wilson (1995). While it may not be very realistic, it does not directly affect the pattern of queue development, which is determined by the overall turning movements as opposed to the routes followed by individual vehicles.

Finally, stochastic variation in the level of demand is allowed. Generation of vehicles from the sources is Poisson with mean demand at each access point μ. A fixed value of μ is used to create an isotropic flow of vehicles through the system as a whole.

The program is intended to deal with large networks at high speed, and care has been taken to minimize the data storage requirements and to streamline the coding. The development of an area-wide traffic jam can be observed in much less than real time, allowing a range of parameter variations to be explored at a reasonable cost in terms of computer resources. The program can handle a jam extending over a grid of 19 × 19 blocks (i.e. 760 links and 400 junctions) in about 2 min, with potential for further expansion.

The existing software includes facilities for manipulating the time interval between successive updates of the vehicle queues, the traffic demands and the queue storage arrangements. Although not specifically intended to do so, the process of switching priority alternately between E–W and N–S traffic echoes the operation of a co-ordinated signal system, and in future it might be worthwhile to develop this aspect to provide a facility for investigating the effects of varying signal timings on jam behaviour. In addition, the software can be used to investigate the way in which jams disperse by removing the blockage and introducing external counter-measures. The graphical display enables the user to view the success of intervention over a considerable time period.

References

Abbess, C.R., Roberg, P., 1995. Fractal structure of traffic jam images. In: Titterington, D.M. (Ed.), Proceedings of the IMA Conference on Complex Stochastic Systems in Engineering. Oxford University Press, Oxford, UK, Vol. 54, pp. 251–281.

Batty, M., 1990. Cities as fractals: simulating growth and form. In: Crilly, A.J., Jones, D.H., Earnshaw, R. (Eds.), Fractals and Chaos. Springer, New York, pp. 43–69.

Holden, H., Risebro, N.H., 1995. A mathematical model of traffic flow on a network of unidirectional roads SIAM Journal of Mathematical Analysis 26 (4), 999–1017.

Huddart, K.W., Wright, C.C., 1989. Catastrophic congestion, and some possible ways of preventing it. TRAFFEX 89 Conference: Seminar on Congestion Control and Parking Enforcement, Brighton, April 1989, PTRC (UK) Ltd (unpublished).

Roberg, P., Abbess, C.R., 1995. Diagnosis and treatment of congestion problems in central urban areas European Journal of Operational Research 104 (1), 218–230.

Wilson, R.E., 1995. Mathematical modelling of road traffic networks. PhD thesis, University of Oxford, to appear.

Witten, T.A., Sander, L.M., 1981. Diffusion limited aggregation: a kinetic critical phenomenon Phys. Rev. Lett. 47, 1400–1403.

Wright, C.C., Huddart, K.W., 1989. Breaking the grip of gridlock Surveyor 172 (5052), 16–17.

Wright, C.C., Huddart, K.W., 1989b. Strategies for urban traffic control. Transport and Society Discussion Paper 6, Rees Jeffreys Road Fund (unpublished).