The Yateley Society

TRAFFIC SPEEDS, FLOWS, GRIDLOCK AND ACCESSIBILITY - A LAYMAN'S INTRODUCTION

Dr Richard H Johnston, The Yateley Society
(formerly Principal Scientific Officer, Transport Research Laboratory)

1. Introduction

1.1 Traffic issues are often raised as objections or concerns in connection with major new planning developments. These questions are often regarded by the professional experts as being purely technical issues which can be overcome by making local engineering improvements near the site. The larger question of the impact on the transport infrastructure generally are normally ignored. This is because it is assumed that the effects further away from the site will be small and negligible.

1.2 However, local people and their political representatives are often convinced that the impact on their lives will be more substantial. They often feel unable to address the issues because the traffic profession speaks in terms of the various standards and memoranda which are produced by the Department responsible for Transport (now DETR, Department for the Environment, Transport and the Regions).

1.3 The object of this paper is to provide some insight into the underlying issues, and to explain, in layman's terms, why traffic behaves as it does, why congestion has some very peculiar characteristics, and why it does not take much extra traffic to turn smoothly flowing traffic into a linear car park.

1.4 I was directly involved for over ten years on transport research for the Government comparing land-use transport interactions in Britain and foreign countries, the viability of public transport, peak-hour traveller behaviour and its implications for congestion, suburban traffic congestion and, later was directly responsible for research on speed-flow relationships on rural roads and motorways, and achieved national level status as a research expert on these subjects. As such I have a wider view of transport and traffic issues than many practicing traffic engineers and planners, but claim no specialist knowledge of the detailed traffic engineering aspects.

2. Some background: DETR Technical Memoranda and the general state of knowledge

2.1 The DETR Technical memoranda and standards are based ultimately upon research observations taken of real traffic on roads in Britain. Observations are taken of the speeds and flows achieved through selected samples of roads and junctions, and generalisations from these are then made in terms of such factors as width of the road and the proportion of goods vehicles. These standards are reviewed from time to time, new sets of measurements are taken, and the standards are revised in the light of the results obtained. I was directly responsible for some of the research which led to an update of the standards for rural roads and motorways in the early 1990s.

2.2 As far as "rural roads and motorways" are concerned these standards are developed by considering roads and junctions in isolation. The interactions between the roads and junctions are not considered. For densely packed urban roads, the interaction between junctions becomes an important consideration, and traffic assessment then has to be considered on a network basis. This is far more complex, and consequently costly, so network assessments are only carried out when absolutely essential.

2.3 In outer suburban areas it is therefore common to use the "rural" approach, and hope that the impacts on more distant roads and junctions will not be significant. However this is increasingly inappropriate: although the suburban road network is not as dense as urban networks, so that obvious interaction between junctions is less apparent, the interactions can be far more significant than they might appear. In my opinion, traffic engineers therefore ought to take much more account of the effect of development on the network as a whole than they are presently accustomed to do in suburbia, and should be adopting the kinds of techniques already in use for urban traffic.

2.4 There is also a need for basic research into the interaction between traffic, transport and different kinds of land use layouts in suburban areas. Most outer suburban areas have developed in much the same way. Hence such research is exceptionally difficult to do since different layouts cannot be compared, and so very expensive to undertake, and the outcomes are so uncertain that little, if any, such research has been done. However scattered patterns of mixed development are likely to be unsustainable, because they encourage the kinds of traffic cross-flows that tend to lead to "gridlock".

3. Traffic flow along road links

3.1 The main constraint on traffic flow in a network are the junctions. However before the reasons for this can be explained, it is necessary to understand traffic flow along a highway.

3.2 Highway engineers think in terms of ***traffic flow***, that is the number of vehicles[1] travelling past a given point during a fixed period. Because traffic flow is uneven, and varies from minute to minute, the shorter the period chosen for the measurement, the more variable is the measured flow, but the flows on a busy road are reasonably steady if the fixed period is 10 or 15 minutes. The flow is therefore expressed in terms of "vehicles/hour measured over a 15 minute period". On multi-laned roads, such as motorways, the flow is usually further expressed as "per lane".

3.3 On motorways, measured over long periods, traffic typically achieves 2200 vehicles per hour or more during busy periods. The highest flows per lane on motorways, as measured in the early 1990s, were reaching around 2700 vehicles/hour per lane over 10 minute intervals at peak periods. Such exceptionally high rates cannot be sustained for long periods - such a flow represents ***average*** headways of only 1.33 seconds. The Highway Code recommends drivers to maintain a distance from the car in front equivalent to 2 seconds, which corresponds to a flow of only 1800 vehicles per hour, if they all had exactly 2 second headways. In reality, even at 1800 vehicles per hour, there is a lot of variation in the headways and more than half the vehicles have headways

[1] A large vehicle like a truck or a bus has more effect on the traffic than a car, so their effect is adjusted by counting each one as if it were, say, 2.5 or 3 cars (the precise figures vary), and hence reexpressing the flow in "passenger car units" (pcus). This additional complexity makes no essential difference to the argument.

which the Highway Code regards as dangerous.

3.4 Traffic speed on motorways is related to flow in a rather complex way. At low flows, the traffic speeds are high, and the speed falls only slowly at first as the flow increases towards 2000 vehicles per hour. Thereafter the speed falls more quickly, but still tends to stay at around 50 mph. But if for some reason the flow is disturbed in some way, and the speed falls as a result to 30mph, the achievable flow will fall by 10-15%. This generates a moving and rapidly lengthening "queue" on the motorway, since the traffic behind is still arriving at the previous rate. The queue length will grow at a rate determined by the difference in the new lower flow rate and the previous high rate. Clearly this means that very high traffic flows (above 2200 vehicles per hour per lane on motorways) are unstable.

3.5 Traffic engineers do not normally look at highway flows in the following way, but it is highly instructive to consider these effects in terms of ***road occupancy*** - that is, how much length of road each vehicle is using at a given rate of flow. The following discussion is expressed in terms of yards and miles because it gives simple numbers. There are 1760 yards in a mile, and so for a flow of 1760 vehicles per hour per lane (close to the typical capacity flow per lane for a rural non-motorway road of good standard), if the traffic is travelling at exactly 60 miles per hour, those 1760 vehicles can be imagined as spread over 60 miles of road, so each vehicle "occupies" exactly 60 yards of road. The same flow 1760 veh/hr travelling at 40 mph would have each vehicle occupying 40 yards of road. Hence, generally, at 1760 vehicles/hour each vehicle occupies as many yards as its speed in mph. (The average time between the vehicles would of course remain the same at slightly over 2 seconds.) But in reality, drivers do not reduce their distance from the vehicle in front in proportion to their travel speed - in fact the distance between vehicles is not much less at 40 mph than what it is at 60 mph. This explains why the motorway lane capacity falls when the speeds fall. Indeed some drivers ***increase*** their headway ***distance*** when the speed becomes less stable because the speeds become more variable and they increase their safety space.

3.6 Moreover the above example reveals that achievable flows fall dramatically if the traffic speed falls really low. Naturally if the traffic speed is zero - a solid traffic jam where nothing moves - the flow becomes zero. At a mean speed of 5 mph, to achieve the above 1760 vehs/hr, each vehicle can only occupy 5 yards - the length of the vehicle without any space at the front - literally bumper to bumper. Thus at any speed less than 5mph a flow of 1760 veh/hr is absolutely impossible, and is practically impossible for average speeds below about 10mph. ***Thus very low average speeds represent very low effective road capacity***, which is why the transition between a road having a good flow of speeding vehicles and being a linear car park can be so abrupt with just a very little more traffic demand or following a slight disturbance to the flow.

3.7 The above explains some of the great mysteries of motorway driving.

3.8 The first is the experience of crawling along in a motorway queue at perhaps 5 mph for some miles, then, suddenly, there is an almost completely empty motorway in front. There is apparently no reason for all this frustrating delay, and you accelerate away to 70 mph. The original cause of the problem, which formed the queue has long since gone. But at 5mph, the motorway can only achieve a flow of perhaps 900 vehicles per hour per lane, but at 70 there is room for 3 times that flow, but the vehicles at the front of the queue have to accelerate away to try to fill the empty space. The front of the queue moves back along the motorway as they leave from the front at more than 900 vehicles per hour, which is why there is the sudden transition, and no obvious cause.

3.9 The second mystery is the rapid changes in speed which can occur, especially in the outside lane of a motorway (which always has higher flows than the inner lanes). Here the close headways and lane changing make flow disturbance very likely, and the high flow can be sustained some of the time. When a disturbance occurs, the flow at that point drops, but the vehicles are still arriving from behind at a higher flow, so a queue forms and the speed there drops sharply. The last vehicles in front of the disturbance are still speeding away, so a gap opens up that the slowed traffic then accelerates into, and the traffic can speed up again, but the queue moves backwards like a shockwave through the traffic behind. These waves representing alternately fast and slow traffic can be seen routinely on all very busy motorways. They are very dangerous, because the amplitude of the speed changes tends to increase as the wave moves backwards, and may lead to accidents. Their effect can be reduced by imposing upper speed limits, which damp out the acceleration phase of the phenomenon by reducing the peak capacity (which is higher at higher speeds).

3.10 In this section we have seen that the vehicle flows which can be achieved on a road depends to some extent on the speed of the traffic - at very low speeds flows are necessarily very low, well below 1000 vehicles per hour, and that there is still some increase in flow in practice as speeds continue rise, though the stability of the flow deteriorates above about 50mph.

3.11 This explains why a few extra vehicles can have a hugely disproportionate effect on traffic flow. Although it is a statement of the obvious, if the flow of vehicles arriving at a point exceeds the flow leaving it, then a queue forms. Stationary queues mean very low average speeds, and hence very low flows, and once a queue starts it therefore grows very rapidly.

4. Junctions

4.1 At a simple cross roads, 4 lanes of traffic are compressed into effectively (at most) 2 lanes through the junction. That in itself is a major reduction in capacity. Also, the crossing traffic interferes and so vehicles must reduce speed in order to travel through that restricted capacity safely. But as already shown, low speeds reduce the capacity still further compared with the flow on the approach roads, and queues will form long before the approach roads are even busy. This is why junctions are usually the main limitation on flows in road networks.

4.2 Of course, better junction designs are possible which have more capacity than a simple junction. Roundabouts improve the stability of the flow, and hence raise the speed through junction and increase its capacity. Two lanes round a roundabout increases it further. Roundabouts with 3 lanes have been tried, but they are difficult to use safely, and the further capacity increase is usually small as a result.

4.3 Traffic lights produce more capacity than a corresponding roundabout. This is a surprise, since all the traffic is completely stopped some of the time. This works because when the traffic is moving, it is moving more quickly and without disturbance, and so achieves much higher flows while it is moving, as explained previously. Extra lanes are sometimes provided on the exits from traffic lights (and roundabouts) - exploiting the fact that once traffic has accelerated speeds are higher less road space will be needed to achieve a given traffic flow.

4.4 One of the problems with junctions of any kind is that their capacity is almost always less than the feeder roads. Queues at junctions are therefore a normal feature, and there has to be enough space to accommodate the queues which are likely to develop, without having an impact on the rest of the network.

5. Road Networks

5.1 Road Networks consist of road links between junctions. If a queue at one junction extends back along a link to the next junction it will block that second junction, and the latter will then cease to function normally. That is likely to cause queuing to form at that junction which then goes on to lock further junctions, and this can cause a whole wide area to be come locked with traffic which is not able to move at all, or if it does so only very slowly (sometimes called gridlock, see appendix for further discussion). Hence a network which with good traffic speeds was capable of handling high flows of traffic has become a network capable of moving only a small fraction of its normal capacity. It is self-evident that this is an eventuality which should be avoided. It is important to note, however, that these conditions can be created by an apparently minor incident (an accident, or broken down vehicle) at just one junction. Anything which reduces the capacity - even for quite a short period of 5 or 10 minutes - can have traffic implications which may take hours to clear if the network is operating close to its capacity.

6. Effect of driver behaviour

6.1 As compared with the off-peak, drivers who regularly drive at peak periods tend to drive more quickly and accelerate more quickly when they get the opportunity. This compensates to some extent for the reductions in capacity available, but the very high flows now achieved on motorways show such improvements are now almost completely used up, and are eating dangerously into safety margins. It is difficult to see how any further increase in flows can be achieved using conventional vehicles driven by their occupants.

6.2 “Science-fiction” ideas for computer controlled road-trains of vehicles seem unlikely to be the answer, even on motorways: public sensitivities imply that the safety space needed means that the flows are unlikely to be much greater than are already being achieved. And high tech solutions of this sort provide no solution for congested urban or suburban networks.

7. Congestion and accessibility

7.1 Suppose traffic demand is increasing, either through the general growth in traffic, or because of continuing new development. In either case, the increase in flow will cause a fall in speeds. When average speeds on a network fall very low, the achievable flows are reduced, as previously explained. The fall in achieved flows lengthens queues and the speeds fall still further in a vicious circle. If people continue to try to make the journeys the time it takes them will get longer and longer, until making the journey becomes effectively impossible. Before this happens, some people will decide not to make the journey: they will change the time of travel, go somewhere else, or take a different mode of transport. The lower bound on speeds will depend on how many such options are available to them, and how good those alternatives actually are.

7.2 A truism, perhaps, but people normally only make journeys to connect the activities they want to undertake. They need to be able to access those places within a reasonable time. The ***average*** person spends about an hour a day travelling, a figure which varies little from place to place. If they live where activities are spread out, as in suburbia, then they need faster modes of travel, like car, than they do where everything is closer together. Thus it is the ***time*** a journey takes which determines whether it acceptable. Walking is obviously slow, but so is public transport (except for journeys over 20 miles) because of the time needed to walk to it, wait for it, and the stops made to accommodate other passengers. The ***accessibility*** of a place is good if a large number of activities are available within a reasonable time of travel, and poor if few places are readily accessible.

7.3 In major conurbations like London, where walking is often possible, and off-road public transport (rail and underground) is good, and there are many alternative destinations, the traffic speeds have stayed fairly high - at about 12 mph, a figure which has changed little over the last century. The average door to door speeds by people using public and private transport in London are remarkably similar, large numbers use each mode, and this suggests that there is genuine competition between the different modes.

7.4 In low density suburban areas, local facilities are often weak, so that most activities require a journey of at least five miles. Public transport, such as it is, must share the roads with other traffic, and so is at least as slow as travel by car. When walking and waiting times and the time lost at bus stops are added, public transport is not a realistic competitor with car travel. In these circumstances, people have no choice but to travel to other places by car, and travel speeds will stabilise at a far lower level than in Central London. If the congestion is such that those 5 miles take 30 minutes to an hour to achieve - as is increasingly the case in North east Hampshire and Berkshire in peak periods, not only is the network capacity reduced by the low speeds, but people are increasingly imprisoned in their homes. Their accessibility to desired destinations for work, shopping or leisure is low.

7.5 Although providing shopping, work and entertainment opportunities near their homes seems the answer, this is not particularly practicable without a very considerable (and unattainable) increase in the overall density of development to urban densities, and with the correspondingly urban infrastructure added. At present those facilities are not being provided either because they are not economic, in the case of shopping or entertainment (not enough demand), or in the case of workplaces, because workplaces require specialised skills which can only be found across a much wider geographical area. In such cases, firms will choose to locate where communications are very good, so that workers can be drawn from a wide area. The poor accessibility in the low density suburbs militates against these developments, and the attempt to add them adds further to the congestion problem and makes accessibility still worse. Activities which generate large numbers of journeys need to be located in places where those journeys can be made easily, not in areas where there is already a problem of limited and road based transport capacity.

8. Conclusion

1.1 What matters to people is that they have good accessibility to the activities they wish to undertake whether that is travel to work, school, shops or entertainments. Their preferred mode of travel will depend on the distances they have to travel to reach those activities. Because the average person only travels for about an hour a day, walking, cycling and bus are normally suitable only if those activities are within 2 to 3 miles of home. Car is essential for making longer journeys (including most rail journeys in suburbia). This explains why car ownership and use is exceptionally high in outer suburban areas; land use densities are generally low, and activities are therefore widely separated, necessitating long journeys. If traffic congestion as a result of high demand flows results in lower travel speeds, this reduces the effective capacity of the network (since flow falls with speed), and results in falling accessibility, effectively trapping people in their homes. The answer to this is not to try to increase facilities in those suburbs, because those facilities will draw from other places and further increase local congestion on the fixed road network. Rather it is preferable to locate facilities in locations which are, or can be made to be, well served by the transport network from all the places from which travellers are likely to come.

Appendix: Gridlock

A.1 Gridlock is an American expression for the conditions which occur when all the traffic in an extended area comes to a complete standstill. This was first reported as occurring in some American cities with grid-pattern road systems as early as the 1950s. It develops quickly but can take many hours for such clogged conditions to clear. Grids are naturally very prone to interconnected locking of junctions since each square has its own complete circuit of stopped traffic, but the same phenomenon can occur on any type of road network. The Americans started to report such gridlock conditions in ***suburban*** areas during the 1980s, so it was only a matter of time before such conditions would start to occur in Britain. Preliminary research I carried out at TRRL in the late 1980s using population, road network density and car ownership data showed that because their ratio of cars owned to road space was very high, the outer Home counties such as Berkshire, Buckinghamshire and parts of Hampshire and Surrey were likely to be the places where suburban gridlock would occur first in Britain. It is now, 1999, happening fairly frequently.

A.2 City gridlock has proved less of a problem in Britain than in America because of a number of factors, such as higher land use development density, better urban traffic control on a more hierarchical road network, and much greater use of non-road public transport (i.e. rail). But these advantages do not apply to Britain's suburban areas, which are closer in form to the American model, with relatively low density development, and dispersed land use activities which demand car use to service them.

A.3 The apparently obvious way to try to overcome a gridlock problem is to install more capacity, but this may be counterproductive. Extra roads imply more junctions, and hence more junction delays, and reduce the possible queue lengths before the next junction back is blocked. In fact, the capacity of a network can actually rise substantially if some links and junctions are removed, as has been done extensively in central London.

A.4 In suburbia, the scope for substantial improvements to road capacity is often very limited, and the main road system in such areas is often little different from what it was in the days of the horse and cart. Environmental considerations, and existing buildings preclude the possibility of making major enhancements. The answer has therefore to lie in more appropriate land use planning. Unfortunately little research has been undertaken to discover what land use planning approaches really do work.

A.5 However it is clear that what must be avoided is criss-crossed traffic flows. The less that traffic interacts with other crossing traffic the better. Unfortunately most British suburban development seems to be laid out in a way which encourages such cross-flows, since suburban land uses are not laid out to encourage radial transport patterns, which tend to minimise the interactions.

A.6 Effective public transport provision demands large numbers of people to want to make the same journey: this is easier with radial journey patterns, and is the key to the success of the London rail commuter system. Public transport cannot be economically provided (and buses cause congestion) if the patterns of travel are diffuse and criss-crossed as they are in the typical British suburban area.