When shock waves hit traffic: What turns a fast-moving stream of cars into a stagnant pool of frustrated motorists? Bob Holmes talks to the traffic engineers who are beginning to crack the gridlock

25 June 1994

BOB HOLMES

You're cruising down a crowded motorway when, suddenly, traffic slows nearly to a standstill. A few moments later, again for no apparent reason, everyone speeds up again. You've passed no slip road pouring cars onto the highway, no crumpled wreck on the roadside, no clue to the cause of the problem. What happened? Puzzling as this common experience is for most motorists, it has a simple explanation in the little-known world of traffic flow theory.

Traffic engineers have known for years that traffic behaves much like a moving fluid, transmitting shock waves of congestion far upstream - and sometimes downstream - from bottlenecks. Now they are beginning to understand better how these shock waves can result from the most trifling causes, and how driving habits can contribute to the disruption. Other kinds of mathematical analysis tell engineers how to time city traffic lights to minimise delays to traffic.

As planners turn their attention from single, isolated roads to the complex warp and weft of traffic on a city's network of streets, mathematical analysis has to give way to computer simulations. With the aid of such simulations, researchers have begun to develop a general theory of traffic flow in complex street networks. But investigators cannot agree on how to study the intractable problem of traffic congestion.

Traffic flow is simplest on multilane roads such as motorways, where everyone is going the same way, and there are no traffic lights or crossing traffic to confuse matters. Here, theorists use three variables to describe traffic behaviour: vehicle density (the number of cars per mile of road); vehicle speed; and traffic volume (the number of cars passing a fixed point per hour), which is equal to the speed multiplied by the vehicle density.

MOTORWAY MANNERS

When the traffic is light, cars do not interfere with one another. Each driver can choose his or her own speed, and overtake slower vehicles at will. Under these conditions, increasing the vehicle density by a few cars per mile of road makes no significant difference to drivers' speeds. The extra cars merely boost the traffic volume in proportion to their number. But as more and more cars enter the motorway, drivers coming up behind a slower vehicle begin to find themselves boxed in by the traffic in neighbouring lanes, and cannot overtake until they are clear. This delay slightly reduces average speeds, with the result that the volume may level off. Eventually, traffic reaches a maximum volume, usually about 2200 vehicles per lane per hour, which engineers call the capacity of the roadway.

Once traffic volume has reached capacity, cramming still more cars onto the motorway slows everyone down sharply - so sharply, in fact, that the volume of traffic flowing past any given point will decrease, even though the cars are spaced more closely. Traffic engineers call this sudden drop in speed a 'breakdown' of the flow, and admit that they don't understand why it happens at some times and places and not at others where the flow is equally heavy. If things get really bad, cars end up in a long queue, creeping forward in a stop-go fashion: in short, a classic traffic jam, and the cause of frayed tempers and burnt dinners for thousands of hapless commuters.

When traffic flow breaks down, it generates what the experts term a shock wave, by analogy with...
compression waves in a fluid. Watching from above, you see a wave of stopping traffic sweeping backwards up the motorway as driver after driver hits the brake and joins the queue. Later, as the jam unclogs, a similar wave of accelerating traffic similarly spreads backwards through the queue.

These moving shock waves are the reason why motorway traffic often stops at places where there is no obstacle to be seen. Think of the similar case of sound waves. You hear the bang of a firework wherever the sound wave reaches you, not when you pass the site of the explosion. Similarly, your car stops on the motorway wherever it meets the shock wave, which may be far from the broken-down car or burst of traffic from a slip road that caused the problem.

Most flow breakdowns happen for fairly obvious reasons, says Fred Hall of McMaster University in Hamilton, Ontario. What drivers have to realise, he notes, is that the cause may lie far ahead of them. Usually, traffic backs up because of a bottleneck caused by a slip road feeding traffic onto an already crowded road, or perhaps by an accident or road repairs.

Now and then, however, breakdowns can happen spontaneously, caused by nothing more than clumsy driving. In closely spaced traffic, drivers worry about hitting the car in front of them, and if it slows slightly - to round a curve, for example, or because the driver is gawping at an accident or just daydreaming for a moment - the driver behind may play safe and brake more than is strictly necessary. That can cause the third driver in line to slow down even more, and so on, in an amplifying wave of deceleration. If enough drivers overreact, those at the back of the queue could come to a halt, even though the driver who started the chain reaction barely slowed at all.

HEAVY RIGHT FOOT

A mathematical analysis conducted more than thirty years ago at the General Motors Research Laboratories at Warren in Michigan by Robert Herman, whom many regard as the father of traffic science, reveals what it is about drivers' behaviour that causes such pointless fluctuations. Herman devised a series of equations that describe how drivers speed up and slow down to maintain their spacing with the car in front of them. Whether one driver's braking is amplified by each successive driver should depend, he predicted, on how quickly drivers respond to what happens ahead of them and how hard they brake. Drivers who anticipate well and change speed gradually can use the gap between cars as a shock absorber to damp out small changes in speed; those who react slowly or have a heavy right foot lose this advantage - though it is the cars behind them that pay the price. 'Smooth traffic is good traffic. The flow is better, and it's safer,' says Herman, who has now retired from the University of Texas in Austin, where he moved after his spell in industry.

Herman road-tested his analysis by measuring how long drivers took to respond to changes in the flow of traffic ahead of them. Car drivers, he found, teetered at the brink of stability, with about half driving well enough to accommodate the lead car's speed changes without amplifying them. Professionals did much better: every one of the bus drivers Herman tested drove well within the margin of stability.

The way drivers maintain the space between their car and the one in front can cause traffic flow to break down somewhere totally unexpected - a mile or so downstream from where a slip road joins, according to a study, as yet unpublished, by traffic researcher Michael Cassidy of the University of California at Berkeley. Drivers hate to let cars cut in front of them, Cassidy says, so those already on the motorway tend to bunch up more tightly to discourage drivers on the slip road from moving in. Once past the slip road, these drivers relax and back off to a more comfortable distance behind the car in front. This deceleration can trigger a flow breakdown, typically about a mile past the merge - an effect traffic experts have often overlooked, Cassidy notes. 'What's quite amazing about this is that despite the fact that merges are the most common problem, all the presumptions on how traffic behaves have been completely misunderstood.' The dynamics are quite different from what people thought, he says.

WAITING ROOM

With the exception of Cassidy's work, however, efforts to understand the causes of flow breakdown and the stop-go behaviour of traffic are dismissed by Gordon Newell, another traffic theorist at Berkeley and, like Herman, a physicist who became interested in traffic flow in the 1950s. All anyone needs to know to understand motorway flow, claims Newell, is where the bottlenecks are and how fast cars pass through them. He likens getting through a bottleneck to sitting in a waiting room: 'It really doesn't make any difference how you move around in the waiting room.'

Newell's theory, which he published last year in Transportation Research (vol 27B, p 281), treats motorway traffic much like a stream of water being poured through a set of nested funnels. Each funnel collects the water flowing through the previous funnel, plus any added water (traffic joining a motorway), less any water that is siphoned off (traffic leaving a motorway). Each funnel's output is simply its net input or the funnel's maximum flow rate, whichever is less. Any excess water simply stays in the funnel until it has room to pass through.

Newell claims that this simple model allows him to analyse flow on a short section of road using just pencil and paper. He ignores how one driver follows another, the spacing between vehicles, the fluctuating speed in a flow, and all the other arcana of most traffic models, and so is able to model the entire freeway system of the Bay Area around San Francisco on a desktop computer. Most other simulations of such a road network require much more powerful computers.
Concise as Newell’s formulation may be, some traffic practitioners say they don’t find it very useful. ‘Newell must be asked a whole lot different questions than I usually get asked,’ suggests Rick Donnelly, a transport consultant based in Albuquerque, New Mexico. Down in the trenches of the war against traffic congestion, he says, planners need to make tactical decisions such as whether to install traffic lights to regulate the flow of traffic entering a motorway and, if so, what rate of switching is most efficient. Questions like these demand more detailed analyses, Donnelly insists.

Once a driver turns off the motorway and onto city streets, traffic planners’ most obvious role - and the one that probably provokes the most complaints - is to coordinate the timing of traffic lights to keep delays to a minimum. To do this, the engineers must juggle the duration of each complete cycle, from green to red and back to green again, and the synchronisation of a series of lights, so that as many drivers as possible hit a succession of greens. Paradoxically for the planners, the most rational strategies for doing this seem perfectly designed to irritate rush-hour drivers.

Take the question of cycle length. Every time the lights change, there is a period of a few seconds when no one is using the intersection; one traffic stream has stopped but the next one has not yet started. If the signal changes often, this wasted time amounts to a relatively large fraction of the total time, which reduces the total flow through the intersection. Signals that change less frequently have a greater traffic capacity, but cars that miss the green have to wait longer before they can move again.

Planners must balance these two demands. According to Nathan Gartner, a traffic-light expert at the University of Massachusetts at Lowell, they tend to favour long cycles for high capacity at busy intersections and shorter cycles to reduce waiting time where traffic is lighter. However, this solution happens to be the one that leaves commuters, who travel at the busiest times of day, with the most time to grumble as they sit at red lights.

On streets with a succession of traffic lights, synchronisation of the entire chain can make an enormous difference to the traffic flow. Deep down, most motorists harbour the conviction that each set of lights should turn green as they approach, allowing them to sweep down the street without slowing. Such a scheme could indeed be designed for a lucky few, says Gartner, but at some cost to everyone else. ‘That works well if you coordinate lights in only one direction, but if you also want to do it in the other direction, you will have to compromise and will not obtain such good coordination,’ he says.

SEEING RED

To minimise the delay to drivers in both directions down a particular street, the best synchronisation scheme is for all the lights to change at the same time, says Gartner. A little thought shows why, when you consider just two junctions a short distance apart. The optimum solution is for both sets of lights to turn green at the same time. Delaying one set by a few seconds will make things better for the traffic moving in one direction, but at the cost of hindering those going in the opposite direction. However, the opposite applies for two-way flow through lights that are spaced too widely for drivers to have much chance of making both lights on a single green. Here the best compromise is for one to turn green just as the next one turns red.

Few traffic-light systems follow this simple rule, however, because reality intrudes, and knocks the symmetry of this idealised model askew. Traffic flow will not usually be equal in the two directions, and planners may want to favour flow in the busier direction. Also, flows may well differ at every junction, as traffic joins and leaves the street. Pedestrians, filter lanes for turning traffic, crossing traffic and other complications such as tram lines only worsen the problem. ‘When you get a fairly large system that is operating under different peaking characteristics at different places, and you’re trying to maintain a steady flow in more than one direction, optimising becomes a very difficult mathematical problem,’ says Donnelly.

Complexities like these have led some traffic theorists, and the vast majority of planners, to downplay mathematical analysis and turn instead to computer simulations of traffic flow in road networks. Such simulations allow researchers to juggle myriad details about the behaviour of drivers and their cars, including the way one vehicle follows another, lane changing and the speed preferences of individual drivers. Theoreticians often describe these different behaviours using totally different mathematical forms, Donnelly notes, which makes a pencil-and-paper solution almost impossible. ‘But if you can see the result of the simulation,’ he says, ‘you start to get an appreciation.’

For designers of traffic systems, simulations offer another precious advantage over reality: the chance to experiment. ‘One of the neat things about simulation is that you can make a change to a problem spot and look for unanticipated changes to the surrounding system,’ says Donnelly. For example, traffic lights to limit the flow of vehicles onto a city freeway might cause a queue to back up into an intersection on a surface street. ‘Our profession is characterised by solving one problem and creating other problems. The beauty of simulation is you can avoid doing that.’

While most traffic experts use simulations to solve problems about specific street layouts, a few are chasing more elusive generalities about the way traffic behaves in any complex urban network. In the late 1980s, Hani Mahmassani and colleagues at the University of Texas at Austin used computers to mimic simple, idealised traffic networks and the larger, messier reality of downtown Austin's streets. As they watched what happened to traffic flow when they added more vehicles, or modified streets
and traffic lights, they were pleased to find the relationships between speed, density and volume that they already knew applied to motorway flow began to emerge for the more complex flow in city streets.

In particular, the simulations show traffic volume increasing as vehicles crowd more densely onto the roads - but only up to a point. As still more cars hit the streets, speeds and flow volumes plummet until the network reaches gridlock - the citywide equivalent of flow breakdown on a motorway. 'It's really a very powerful result in traffic theory,' Mahmassani says. City street networks are too complex for researchers to derive these results mathematically, as they can for a motorway. 'But through simulation, we are able to verify that these relations are similar,' he says.

Mahmassani recognises that there are some important differences, as well as similarities. For example, some streets carry more traffic than others, so densities at the busiest intersections can be much higher than the average for the whole city. As a result, networks start to become congested at much lower average densities - as low as 30 cars per lane per mile averaged over the network, as compared to about 80 for single motorways, Mahmassani says. In short, the traffic capacity of a city is less than the sum of its parts.

The complexity of street networks can also contribute to gridlock directly. 'Gridlock typically occurs when traffic patterns are complicated and conflicting,' says Mahmassani. 'Because you have turns and intersections (with crossing traffic) and so on, things can just lock up completely.' Planners fight gridlock by simplifying traffic patterns - prohibiting turns during rush hours, for example - as well as by trying to divert cars away from the problem area. On the bright side, Mahmassani notes, networks also offer drivers the chance to escape congestion by choosing alternative, less congested routes.

But beyond these crude generalities, Mahmassani says that gridlock, like its more straightforward cousin flow breakdown, remains somewhat mysterious to traffic analysts. Researchers can calculate the flow volumes and densities that cause congestion, but they still know little about exactly where, when or why traffic flow ultimately collapses. And, until they do, traffic planners can never be certain they are following the best strategies to get commuters off the roads and home to supper.

From issue 1931 of New Scientist magazine, 25 June 1994, page 36

Close this window

Printed on Mon Dec 24 05:32:04 GMT 2007