Understanding Road Congestion as an Emergent Property of Traffic Networks

Ed MANLEY
Department of Civil, Environmental and Geomatic Engineering, University College London
London, WC1E 6BT, United Kingdom

and

Tao CHENG
Department of Civil, Environmental and Geomatic Engineering, University College London
London, WC1E 6BT, United Kingdom

ABSTRACT

Despite a considerable amount of research into the modelling of traffic flow through road networks, a clear understanding of the conditions that cause and exacerbate urban road congestion remains elusive. This paper presents a novel approach to this problem, by identifying congestion as an unintentional emergent property of driver-to-driver interactions. It is described how human interaction is a constant and important intrinsic property of city driving, and how these interactions can lead to explicit phenomena. A framework is presented for the further analysis of congestion at three spatial levels, and the driver behaviours that may be of most significance at each described. Building on this model, the paper presents a case study carried out at the highly congested Blackwall Tunnel in London. In providing evidence for some of the concepts described in this paper, the case study demonstrates how human behaviour causes the emergence of road congestion. The concepts and framework presented in this paper represent a strong starting point for further understanding and eventual modelling of the occurrence and spread of road congestion.

Keywords: Urban road congestion; Traffic; Complexity; Driver behaviour; Driver interactions; Multi-scalar.

1. INTRODUCTION

Despite the continuing efforts of traffic engineers and academics alike, the problem of road congestion continues to blight cities across the world. The problem is not only unsightly and inconvenient, emissions from idling cars represent a serious threat to the respiratory health of children [1] and is a crippling cost to the economy (estimated at a cost of least £2 billion per year in London alone [2]). The problem is often assumed to be that there is not enough network capacity – and indeed this is one key constraint placed upon any network – however, in reality, congestion is much more complex than this. Braess’s Paradox and the theory of induced traffic demonstrate that building roads is not always the right strategy. While, perversely, driver’s attempts to avoid congested areas of the network may, in fact, precipitate it elsewhere. According to Bovy and Hoogendoorn [3], congestion is affected by factors of both in macro- and microscopic scale. Route-choice, land-use, demographics, time, fuel price, origin, destination, light conditions, weather, network and geographical topography and driver experience all have an impact on the level of traffic on a given road. But it is the local composition of traffic, the behaviour of drivers and roadway design that actually trigger its development. So, it may be reasonably suggested that it is the interactions of thousands or millions of decisions and events that lead to a congested network. However, unfortunately, beyond this little more is understood of the complexity of the interactions that lead to its formation and propagation around the network. In order to model this, the whole system of macro- and microscopic interactions must be considered.

The study of congestion has conventionally been approached from the perspective of a single scale. At one scale, mathematical assignment models, developed by Wardrop and others, allocate aggregate flows across a given road network. The reasonable assumption being that traffic eventually reaches equilibrium where it ‘arranges itself...in such a way that no individual trip maker can reduce his path costs by switching routes’ (Wardrop’s First Principle [4]). While these models effectively, and elegantly, forecast the flows across a network, they clearly take a macroscopic perspective. They therefore implicitly discount the potential microscopic behaviour that may cause, or prevent, congestion. The same problem is found with land-use models, which use topography and demographic data to predict traffic flows but again ignore local interactions. At the other end of the scale is traffic simulation software. Well known commercial products such as VISSIM, AIMSUN and Paramics seek to represent driver behaviour on the level of an individual driver [5]. However, the emphasis of these models is more upon the representation of actions in ‘normal conditions’, representing only activities such as lane changing and how cars follow each other. Driver behaviour is assumed to be homogeneous, and the irregular or uncooperative actions that sometimes precipitate congestion are not considered. Further, route-choice is generally modelled simply, either using static routing (that is non-adaptive) or guided by macroscopic models, such as those described above. Such approaches ignore the complexity of human decision making on the road network, which may be particularly variable in a congested network.
It is clear that a new perspective to understanding this phenomenon is required. The link between microscopic, individual behaviour and macroscopic, network performance must be better understood. The complete cause (human behaviour) and effect (congestion) relationship must be better understood in order for the situation to be effectively modelled.

This paper will introduce the concept of congestion emerging as a macroscopic effect of human behaviour on the road network. This model will demonstrate that the performance of the road network is inextricably linked to the behaviour of those using it. This paper will therefore examine how the ideas of network complexity and emergence may be used to shed new light on this issue. The next section will begin by examining the exact nature of emergence and its application in other scientific disciplines. Following this, in section three, the apparent presence of emergence within the urban traffic realm is discussed, and in particular how driver behaviour is inextricably linked to it. Section four will move on to detailing more formally how congestion may spread, and the implications this model has for future research. Finally, in section five, a case study will begin to identify the behaviours that cause and encourage congestion at the lower scale levels of interactions. Discussion will then be offered to as to how this new way of understanding congestion may impact on future modelling solutions.

2. EMERGENCE IN COMPLEX SYSTEMS

In order to begin to describe congestion in terms of an emergent property, it is vital that a strong and correct understanding of the term is established. The concept of emergence is, after all, widely used in everyday English and increasingly in scientific literature. In its modern usage, emergence might be most commonly associated with its prevalent use in news reporting to describe the recent or on-going development of stories (‘It is emerging tonight....’, for example). This use is widely understood and accepted, yet is misleading in terms of its usage in this paper.

Emergence may be defined as the temporal process by which the macroscopic properties of a system or network alter due to the microscopic changes of its constituent parts. It is this micro-macro link that is most important here. Emergence is not simply a matter of system evolution; it is an alteration in one or many constituent parts that lead to a new emergent system state. The changes in the system generated from this micro to macro impact may be described as emergent properties. That is to say, the property of change can not be ascribed to any one individual, and may only be viewed as a property of the new system itself.

This concept of emergence has been studied at a number of scales. Philosophical emergence relates to the understanding and epistemological position of a concept or mindset [6]. Systemic emergence (relating to systems of living beings and/or objects) looks at the relationships between objects within that context, and whether the systemic output can be suitably explained by its constituent parts [7]. While network emergence relates to the small temporal changes in mapping function that changes a dynamic network structure over time [8]. In this paper, emergence will be looked at from a systemic perspective. While the systemic viewpoint enables fruitful discussion about the nature of the emergence of congestion, the network complexity toolkit is useful in offering formal definitions about its evolution.

Emergence has been identified in many fields of science and engineering, and the concepts described in these areas can be useful in exploring others.

Natural Sciences

The most well known examples of emergence can be found in the natural world, where the interactions of many creatures can lead to beautiful yet only recently understood phenomena. Bird flocking and ant colonies have both been discovered to be a result of the individual behaviours, whereas previously research focussed on how and who orchestrated the collective.

Social Science

One important area of research into emergence has been towards identifying how the behaviour of many individuals can lead to complex and misunderstood phenomena. Recent developments in sociology and psychology have recognised the importance of understanding the collective at multiple levels. Without a cross-disciplinary approach, emergent crowd behaviours such as mass hysteria and aggression [9] might otherwise be poorly understood. Emergent behaviour of crowds can also be seen in everyday situations. Group actions such as queue and lane formation (in condense crowds) occur through the independent actions of multiple individual. Taking subtle, unspoken hints from what they see around them, people take a decision to act in a certain way – the accumulative result of these actions result in these aggregate emergent properties.

Technology

Recent studies into Internet traffic and network complexity have also noted the apparent emergence of self-organisation and congestion. In the same multi-level way crowd dynamics has been approached above, Smith [10] recently suggested that a similar approach should be taken to understanding these phenomena. The management and prediction of Internet traffic carries many of the same constraints as road network traffic (minus the human complexity), so some useful lessons may be learnt from such analysis. Emergent properties have also been identified as a characteristic poorly understood by traditional (reductionist) software system design methodologies [7].

3. ROAD CONGESTION AS AN EMERGENT PHENOMENON

It is clear from the above examples that emergence can be subtle, surprising, distrusted and yet ultimately vital in explaining hitherto poorly understood systemic phenomena. However, as suggested earlier, there has been little impact made in understanding road congestion in this holistic sense. Previous models have focussed on the capacity-flow problem, based on the ‘normal’ behaviour of individuals or groups of individuals. But these haven’t provided true insight into the problem. Instead, as will be described below, congestion is a consequence of human behaviour and human usage of the road network. While this behaviour is obviously constrained by the nature of the network they use, ultimately it is the individuals that generate the emergent properties of the system.
Congestion is not the behaviour of any single person. No one driver gets into their car and decides they want to cause road congestion, they do not have the power to be able to do so. A driver may decide to park in the middle of a motorway, causing everyone behind them to tailback, but the actual congestion, the system property, only emerges as drivers arrive at this blockage and are unable to pass. The arriving drivers become the blockage for those behind them, and the process – congestion – continues to spread. Equally, if the aforementioned malicious driver was to park on the motorway again yet the other vehicles were able to pass without impact upon flow, then no congestion would form. Congestion doesn’t emerge here despite the potential for it to do so. Like a fire, congestion needs a spark and then fuel – vehicles in this case – to maintain it. In this sense – like emergence – congestion is a temporal process, one that forms then spreads or dies out, and it is the rate at which it moves that is of most interest to drivers and traffic engineers alike.

The spread of congestion must therefore be intrinsically linked to the people that use the roads. At a macroscopic scale, the location of these people, in relation to a ‘congestion cause’, is vital to its impact. Clearly, should a busy route, with many users trying to pass through, become blocked, this would certainly lead to severe tailbacks. These queues would naturally spread to the surrounding network, causing even further delay. But the rate of spread depends on the location of other people around this delay, and the actions that they take. Extending the example above, its spread could be theoretically halted at one in-flow junction were drivers made aware with enough time that their chosen route was blocked. With this information, and providing other route options were available, they’d be able to alter their route, potentially lessen their delay, and decrease the impact of the building congestion.

Indeed, such methods are used widely, usually to advise drivers of long-term road works. At these points, to an extent, the impact of the blocked road and associated congestion are decreased by the actions of the individual. Yet, unfortunately, this adds further complexity. Extending the scope further outwards, another widely recognised phenomenon becomes apparent. The displacement of people from one area of the network, usually well equipped to dealing with large volume of traffic, to another, causes problems in other areas of the network. Large volumes of redirected traffic inevitably lead to capacity issues on smaller roads. So while redirecting drivers around long-term road works may negate one capacity issue, it potentially causes other, possibly greater problems elsewhere in the road network.

It is not only the location and route choices of drivers that can cause congestion to emerge. Recalling the fire metaphor used earlier, congestion needs an initial spark or trigger to prompt its development. Equally important in this analysis therefore are the microscopic individual actions that cause it. However, while it is easier to theorise regarding the importance of location and volume of vehicles on a network to the spread of congestion, identifying the specific actions that cause it is trickier. In normal situations, behaviours including merging cooperation, unexpected or overzealous braking, and lane changes, as well as certain manoeuvres by larger vehicles (buses, lorries etc.) may all potentially, in areas of high traffic density, be the trigger that causes congestion to emerge. At the current time little research has been conducted into the effect of these actions on traffic flow. Further discussion will be offered on this subject during analysis of the case study, described later.

The above discussion describes how congestion is shown to emerge through the actions and interactions of individual drivers. It should be added that the effect is not linear. Congestion may emerge around an incident or area of high flow, but further impact may emerge at other sites far from the initial problem. Drivers can therefore be said to be intrinsically ‘connected’ whilst using the road system. They exist within a network of drivers that is somewhat independent of the road system. The responsibility for the smooth function of the network, while complicated by the topology of the road network, ultimately depends upon the connections between users and the decisions they take.

### 4. TOWARDS A FORMAL MODEL OF THE EMERGENCE OF CONGESTION

With the foundations of systemic emergence of congestion established, there must now be greater focus on how these ideas operate within a constrained, densely packed urban road network. As discussed earlier, congestion may start in a given area for many reasons, but the way it spreads across the network depends on the behaviour of the individuals that make up the network of traffic. It is the way in which it spreads that really begins to effect whole network performance. In order to establish the nature of its emergence, its impact and influence must be examined on multiple network levels. The following section details the theory behind the development of a model for the emergence of congestion.

**Link level**

The link described here relates to a stretch of road relatively unaffected by major junctions or intersections. The formation of congestion in these areas is therefore predominantly caused by driver behaviour, as opposed to traffic control-induced delays. Equally and conversely, driver behaviours can help smooth traffic flows in times of high density traffic. Therefore it is an important place to begin analysis.

The case study, to follow in the next section, will help to begin to reveal the exact behaviours that contribute to the emergence of congestion, or contribute to the continuation of free flow traffic. At this point, it may be suspected that individual actions such as braking manoeuvres and lane changes as well as specific actions attributable to the traffic mix (e.g. large vehicles, bicycles, motorcycles etc.). As has been demonstrated by macroscopic studies into traffic flow, such individual behaviours can precipitate waves of delay along the condensed traffic queue [11]. However, despite this understanding, little work has been done to identify the actual local behaviours that contribute to this phenomenon. The case study, detailed in Section Five, will begin to examine the link between driver behaviour and the emergence of congestion.

**Junction level**

Congestion, then, begins to emerge at link level, where vehicle interactions cause delays that spread backwards in wave-like motion through the traffic queue. However, when congestion reaches junction level the effects can not be assumed to be so linear. Here individuals are provided with choices – choices regarding their cooperation with individuals and choices
regarding route choice. In the first instance, junctions are points where individuals naturally interact and rely upon each other (in many instances) to ensure a smooth transition through the junction. The introduction of new lanes of traffic at junctions can cause further complications within an already congested system. The merging and cooperation actions required here to ensure the smooth flow of traffic instead have the potential to further increase delay unequally across the interacting streams of traffic. It is important to seek to understand how an individual’s interaction behaviour can impact upon not only single link congestion, but also congestion on those surrounding links.

Another choice drivers have at junction level is whether to change route. In normal conditions a driver won’t need to think in depth about this, they’ll know roughly the direction to take. On congested roads, however, this becomes an issue. Drivers, particularly those travelling at rush hour, don’t want to spend long waiting in traffic and so, recognising congestion emerging around or in front of them, may decide to take a different route. This is an important decision as it has the potential to impact heavily on the network around them. One driver deciding to take a different route to avoid a congested road will probably not make a great difference, but many cars making the same choice undoubtedly will.

Network level
The diagram above (Fig. 1) indicates the general idea of this model – how congestion may effectively move from link level, through local junctions and then eventually to affecting a wider area of the road network. The causes and decisions that lead to congestion emerging at network level is not at all well understood, with presently no academic research being carried out at this scale. However, it may be reasonably expected that such levels of congestion emerge only through multiple link and junction failures. Such situations might only be expected to impact upon small areas of the network, or wider areas where large, unanticipated volumes of vehicles are on the move (e.g. post-event, evacuation etc.).

The three-scale model presented here provides a formal definition for the emergence of congestion. Congestion does not simply occur then move backwards, its impact is variable across scales. However, in order to understand its potential spread through a network, the decisions that people make surrounding its development must be better understood. This means looking at both local interactions (as will be examined in section five) but also routing decisions (discussed in Future Work). The following diagram (Fig. 2) demonstrates the links that exist between the decisions individuals make at both link and junction level, and their impact on the network. The actual behaviours that may cause congestion are not referred to here, but further indication of the potential factors influencing individual choice is given attention.

The model presented here has potential as a framework on which modelling software may be developed. By beginning to identify the behaviours that occur at each level, which contribute to the spread of congestion, it may be possible to better predict the impact of road events or measures.

5. CASE STUDY

The concepts discussed so far in this paper are at this stage only that – conceptual. In order to demonstrate the applicability of this model within a real-world context a case study was completed.

Location
The location for the case study was the Blackwall Tunnel in the East of London. The tunnel represents a key route across the Thames River, particularly for those travelling north to the City or Docklands areas of London (the city’s key financial districts) from anywhere in the South-East region. Transport for London (TfL), the organisation that controls all forms of transport in London, have a number of monitoring projects operating on ensuring the satisfactory operation of this key through route. This particular work was carried out with the help of one such project, the TfL Streets Tunnels Operations Centre (LSTOC).

The focus area for this case study was the northbound entry to the Blackwall tunnel, on a weekday morning (6am – 10am). It is widely known that the northbound route generally becomes very congested during the weekday morning peak as people arrive for work in the city. During the afternoon peak there is relatively little flow through the northbound tunnel, as most commuters head back from the north to the south of the river.

The exact section of the network being observed during this analysis was the final mile of the tunnel approach. The northbound tunnel is two lanes wide, with significant height restrictions (4 metres on inside lane, 2.8 metres on outside lane). During this final mile before the entrance of the tunnel the road
is divided into three lanes, with two further inflow lanes entering onto this road. There is additionally another inflow lane closer to the tunnel entrance restricted to buses only. Within the last 300 metres before the tunnel entrance the three lanes are merged in to two for final entry into the tunnel.

To carry out observations of this section of the route, six remotely controlled CCTV cameras operated by LSTOC were utilised. These cameras provided good coverage of all lane reductions and merges described above.

Findings
In tune with the concepts laid out during this paper, the key aim of this case study was to observe the emergence of congestion from the interactions of individuals. By beginning observations at 6am, where flow was evidently busy although not near stationary (vehicle flow data was not available for this section of the road network), it was possible to observe how flow visually decreased and congestion emerged through the interactions of individuals. The key observations relating to driver behaviour were as follows:

- The most influential (and negative) driver behaviour was observed at the three merges prior to the entry of the tunnel. At these points a great deal of selfish, and often illegal, behaviour was observed. Examples of the specific actions were observed as below:
  - Drivers used the hard shoulder (reserved for breakdowns) as a quicker route around the main queue.
  - Some drivers used the restricted bus lane merge point to enter onto the main section of road.
  - Some drivers attempted to cross two lanes of traffic at an inflow (causing delay across two lanes).
  - Many drivers did not pay attention to lane merge directions, instead seeking to gain further progress (illegally) and merging abruptly. The video image above (Fig. 3) demonstrates one such occasion where three merge points (circled) developed from a single inflow lane.

- Many merge behaviours were aggressive and reciprocate with a lack of cooperative yield behaviour. The resulting interactions resulted in slower merging behaviours than would otherwise occur and greater delay to those following

- Buses and goods vehicles were noted to be particularly slow in their merge and queuing behaviour. Although this is clearly attributable to their size, the increased delay during the stop-go motions of the traffic jam was clearly significant. The outside lane height restrictions too meant a limited number of goods vehicles were forced to make late, disruptive lane changes.

- Where illegal manoeuvres were observed, a tendency to ‘follow the leader’ was also seen. For example, when one vehicle was seen to use the hard shoulder to gain advantage over others, a number of others would follow. The general concept of ‘if they’re doing it then I might as well too’ was observed on a number of occasions.

It is important to note that at no point were any complete physical blockages or restrictions placed upon the road. Although within this route there are inflow lanes and a decrease in lanes, the merges are well signalled, with long distances to enable – in principle – an efficient integration of the lanes. The final approach into the tunnel, beyond the final merge zones, was observed to be comparatively much quicker throughout the experiment.

Despite these measures, generally speaking, the observed situation was disorganised and chaotic, characterised by selfish and often illegal individual behaviour. The impact of this behaviour appeared to be significant. Instead of smooth gradient merges that would allow both the merging and main lanes to continue at, albeit, slow speed, the irregularity of behaviour caused an increase in stop-start manoeuvres within the traffic stream. The additional merge points, resulting from selfish action, appeared to prevent a quicker return to a reasonable lane speed, as each further interruption led to the launching of additional stop-start waves. The wave-like formations were observed to move backwards through the traffic network for, in some cases, many hundred metres at a time. These observations are in keeping with those described by Sugiyama et al. [11].

The findings of this case study may be symptomatic of what many experience on the road – drivers acting selfishly, cutting in and causing frustration. However, the true extent to which these behaviours are detrimental to the wider network is not apparent until observed at a higher scale. The results clearly show an important relationship between the emergence of congestion and the behaviour of individual drivers.

6. CONCLUSIONS AND FUTURE WORK

This paper has presented a new way of looking at traffic congestion within the urban road network. It has been argued that congestion is not simply a matter of lots of cars, restricted capacity and traffic lights; it emerges from the interactions of individuals. The behaviour of one or more people can lead to temporary (even delay of a couple of seconds can cause significantly greater delay further down stream in busy conditions) or even permanent delay. It is argued that it is these
behaviours that must be examined in order to truly understand congestion and how it forms and moves.

When a driver decides to take a trip through a city, they'll probably hope that the 'traffic is good' as if traffic may be discussed in terms of it being a single entity controlled from some central command. However, traffic is not like that. The new driver has as an important role within the network as anyone else does. Everyone travelling in the same direction behind them has an intrinsic reliance upon their performance, as they have a reliance on those ahead of them. Everyone within the traffic network is connected, and congestion emerges, because of a number of reasons, from these interactions.

This paper has demonstrated, with primary evidence, that the behaviours of individuals can have a clear impact upon the performance of the whole traffic stream. Although the case study used here was relatively limited in its scope, the clear impact of these behaviours and support for the concepts described here have been demonstrated. Perhaps most striking of the conclusions from the Blackwall Tunnel work was that, even though traffic was able to pass very smoothly through the tunnel, miles of congestion were observed downstream from merge points. At these merges, selfish and often illegal behaviours were regularly observed. Each incidence of this slowed even further a traffic stream with already limited momentum, and sent 'waves of delay' travelling backwards along the queues. Although it is apparent that lane merges will always have some delay upon the flow of traffic, it is hypothesised that more considered and cooperative human interactions would result in a significantly quicker throughput. The way in which people may be convinced to behave in a better way remains a question to be answered.

The results of this case study, although very useful, must be now analysed within the broader framework of that discussed in Section Four. Of equal importance to the spread of congestion is how drivers choose the routes they take, and how these choices impact on the wider network. Perhaps the most significant finding might be to identify all of the behaviours that lead to local network gridlock. Congestion is not solely about momentum, and sent 'waves of delay' travelling backwards along the queues. Although it is apparent that lane merges will always have some delay upon the flow of traffic, it is hypothesised that more considered and cooperative human interactions would result in a significantly quicker throughput. The way in which people may be convinced to behave in a better way remains a question to be answered.

The next steps in this work will involve developing an improved understanding of how congestion develops and moves. This will require further observations, with more concentration on the quantification of results, as well as including information from other sources, such as flow and speed data. By using samples taken from different times and conditions, a much greater understanding of the relationship between the behaviours discussed in this paper and overall traffic flow may be garnered. As described in the aforementioned framework, the spread of congestion is also intricately linked to the routes people choose to take. Therefore, another key strand of research will involve investigating how people make these choices, in normal and irregular situations.

By expanding the understanding of these behaviours the key aim will be to develop simulations that can effectively replicate this activity. In doing so it may be possible to better identify the measures that may be taken to more effectively manage and reduce urban road congestion.

ACKNOWLEDGEMENTS

This work is part of the STANDARD project – Understanding Road Congestion in Central London, supported by the UK Engineering and Physical Sciences Research Council (EP/G023212/1) and Transport for London (TfL).

REFERENCES


