Economics of Transportation 4 (2015) 241–255

Contents lists available at ScienceDirect

Economics of Transportation

journal homepage: www.elsevier.com/locate/ecotra

Congestion in the bathtub

Mogens Fosgerau[a,b,*,1]

[a] Technical University of Denmark, Denmark
[b] Royal Institute of Technology, Sweden

ARTICLE INFO

Article history:
Received 24 April 2015
Received in revised form
19 August 2015
Accepted 19 August 2015

Keywords:
Dynamic
Congestion
Urban
Traffic
Bottleneck
Bathtub

ABSTRACT

This paper presents a model of urban traffic congestion that allows for hypercongestion. Hypercongestion has fundamental importance for the costs of congestion and the effect of policies such as road pricing, transit provision and traffic management, treated in the paper. In the simplest version of the model, the unregulated Nash equilibrium is also the social optimum among a wide range of potential outcomes and any reasonable road pricing scheme will be welfare decreasing. Large welfare gains can be achieved through road pricing when there is hypercongestion and travelers are heterogeneous.

Anybody living in a major city will appreciate that congestion is a significant issue for economic policy. For the US, for example, it is estimated that urban road congestion in 2011 caused a total of 5.5 billion hours of delay (Schrank et al., 2012). Congestion is not only costly. It also has impacts on the local economy, it affects the functioning of labor markets, and it is an offsetting force balancing urban agglomeration effects.[2] It is therefore important for a range of economic issues to understand the nature of urban traffic congestion.

Traffic congestion is essentially dynamic: the traffic system has memory and conditions at one point in time affect conditions later on the same day. Therefore the timing of trips is fundamental and must be taken into account by economic analysis. The dynamic aspect of traffic congestion matters also from a spatial economic point of view due to the connection between the timing and the length of commutes (Fosgerau and de Palma, 2012).

The seminal Vickrey (1969) bottleneck model has shaped our intuition about urban congestion dynamics.[3] The bottleneck model allows the inconvenience of the timing of trips as well as the dynamics of congestion to be accounted for in the economic analysis of congestion. The constant capacity of the bottleneck implies that delaying arrivals at the bottleneck can reduce delays; nobody will arrive later, provided the bottleneck capacity remains fully utilized. This feature of bottleneck congestion implies that a time varying toll can be designed to induce drivers in the middle of the peak to delay their departures, such that revenue is raised, queueing is reduced and no driver is made worse off (Arnott et al., 1993).

Turning to flow congestion, it is well established that the instantaneous speed at a single point on a road is a decreasing function of the instantaneous density of cars at that point (Greenshields, 1935). The fundamental identity of traffic flow holds that flow, i.e. the number of cars passing the point per time unit, equals speed times density, where density is the number of cars per distance unit. Flow is then (with appropriate shape restrictions on the speed–density relationship) an inverse u-shaped function of density. On the downward sloping part there is hypercongestion as higher density is associated with both lower flow and reduced speed.

A recent range of contributions have shown that such a fundamental diagram of traffic flow also applies at the level of an urban neighborhood meeting certain conditions (Daganzo, 2007; Geroliminis and Daganzo, 2008; Daganzo et al., 2011). The underlying mechanism is that drivers continuously adapt their route choices to avoid more congested parts of the road network. This adaptation process tends to equalize congestion across space. A stable relationship emerges between the density of cars in the

---

* Correspondence address: Technical University of Denmark, Denmark
E-mail address: mf@transport.dtu.dk

[1] I have benefited from comments by Ken Small, Jan Brueckner, Per Olsson, Jos van Ommeren, Antonio Russo, conference audiences in Atlanta, Copenhagen, Toulouse and Stockholm, as well as support from the Danish Strategic Research Council. I am particularly happy to have received many insightful comments from Richard Arnott.

[2] See, e.g., Duranton and Puga (2004); Rosenthal and Strange (2004), and Moretti (2011).

[3] Vickrey's paper is extensively cited and has spawned a lively literature on regulating congestion dynamics, see de Palma and Fosgerau (2011).

http://dx.doi.org/10.1016/j.ecotra.2015.08.001
2212-0122/© 2015 Elsevier Ltd. All rights reserved.

network and the space-averaged speed. This is a very important finding, since it allows urban congestion to be analyzed in an aggregate manner, without having to refer to specific road networks. I shall refer to this type of congestion as *bathtub congestion*. Just like the water level is the same everywhere in a bathtub, the level of congestion and hence the speed is the same everywhere in an urban area subject to bathtub congestion.[4] Figuratively speaking, we can think of a car that drives a trip of a certain length in a bathtub: it does not matter where it begins and ends its trip, its effect on the speed of other cars depends only on *when* it is present in the bathtub.

The bottleneck model does not describe bathtub congestion well, since the inverse-u relationship between flow and density does not occur in the bottleneck model. Flow out of the bottleneck does increase with density before the bottleneck until the point where the capacity flow is reached. At higher densities, however, the flow does not decrease but stays constant. Thus the bottleneck does not generate hypercongestion.

Bottleneck congestion may be considered appropriate as a description of urban congestion for example concerning commuting flows towards a city center, where congestion is concentrated near the entrance to the center. Given the now existing empirical evidence, bottleneck congestion can no longer be considered appropriate as a description of congestion at the urban level. For homogeneously congested downtown urban areas, we now have empirical evidence that bathtub congestion is an appropriate description.[5]

This paper presents a model that I call the bathtub model. The bathtub model is similar to the bottleneck model in describing a fixed mass of homogeneous drivers who care about the timing of their trips. The main difference is the congestion technology embodied in the model. Where the bottleneck model builds on bottleneck congestion, the bathtub model (unsurprisingly) builds on bathtub congestion. Thus it incorporates hypercongestion, allowing increases in flow to be associated with increases in speed. In this paper I show that the bathtub model can be used to give a unified treatment of a range of issues related to urban congestion and hypercongestion, as discussed in the following.

The bathtub model takes the length of any specific trip to be exogenous and constant. Thus the model does not describe the mechanism mentioned earlier where speeds are equalized across space as drivers adapt their route choice to avoid congestion. The bathtub model takes as fundamental the outcome that the speed is the same everywhere, such that the spatial dimension can be ignored.

The engineering literature considers a second macroscopic relationship called the network exit function (Gonzales and Daganzo, 2012), which relates the rate at which trips are completed to density. I do not employ this relationship as the model presented here produces the time at which trips are completed as a function of departure times and trip lengths using only the macroscopic speed–density relationship. The bathtub model that I present is not consistent with a network exit function. This means that a network exit function must be considered as an approximation and one that is not necessary in the present model.

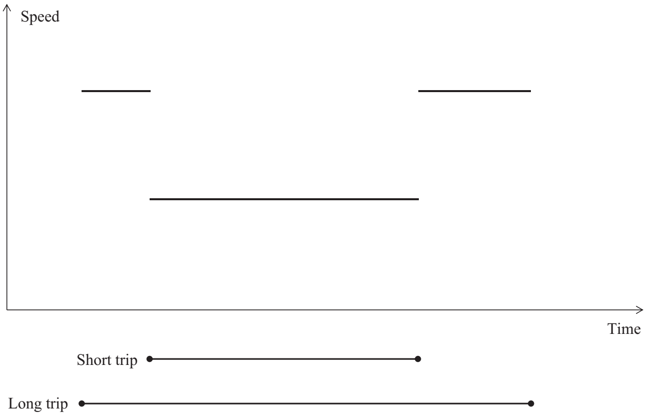

Fig. 1. A short and a long trip with regular sorting. Durations are independent of timing.

The bathtub model leads to conclusions that are radically different from those of the bottleneck model. The bottleneck conclusions depend on the property of bottleneck congestion that it is possible to delay departure times without affecting arrival times. In the bathtub, such trip retiming has small or even no effect on travel times under some circumstances. The underlying principle is illustrated in Fig. 1, which shows a short trip and a long trip in an urban area subject to bathtub congestion. Each trip has some fixed length and a duration that depends on the average speed obtained. The short trip is carried out within the duration of the long trip; I call this *regular sorting*. The speed is low when both trips are ongoing and high when only one trip is in progress.

Notice first that the duration of the short trip does not depend on the timing of that trip. Under regular sorting, the speed for the short trip is always low. Notice next that, still under regular sorting, the duration of the long trip is also independent of the relative timing of the two trips: the long trip covers the same distance as the short trip during the interval when both are ongoing; the remaining distance is covered at the high speed, which is the same before and after the short trip.

Section 2 generalizes this simple example to the case where there is a continuum of drivers with a distribution of trip lengths and shows that travel times for all trip lengths are completely determined under regular sorting. Thus the specific departure and arrival time profiles do not matter at all for travel times. Under a regularity condition, it is shown that Nash equilibrium in the timing of trips is in fact regularly sorted. Moreover, taking the travel time as given, each driver travels at his optimal time. Since travel times cannot be reduced as long as regular sorting is maintained, this implies that the Nash equilibrium is also the social optimum among regularly sorted outcomes. Hence any policy that changes the departure schedule can only make drivers worse off, if regular sorting is maintained.

Section 3 allows demand for car travel to be elastic in two different ways. First, by introducing an alternative mode of travel. I call it "transit" for concreteness, but the defining characteristic is just that it provides a speed that is attractive when car speed drops due to congestion. Travelers have no specific preferences regarding mode of travel. In equilibrium, the availability of transit allows travelers with short trips to escape from the lowest car speeds during the height of the congested peak and instead travel at the higher transit speed. The remaining car drivers are those with trip lengths above some threshold and they gain a speed increase from the absence of the transit users. This mechanism is similar to that described in Anderson (2014), who argues that transit users are

---

[4] Richard Arnott has pointed out that I use the term "bathtub model" in the sense of hydrology and he prefers calling it an isotropic model. Vickrey worked on what he also called a bathtub model of congestion, which was based on the intuition that now materializes in Daganzo's work. Vickrey never completed this work but a note has been preserved (Vickrey, 1991). He used as fundamental the idea that outflow from the bathtub is proportional to the height of the water in the bathtub. This is similar to the second macroscopic relationship discussed below whereby the rate of trip completion depends on density. My model simply computes the times when trips are completed as a function of departure time and speed.

[5] Ji and Geroliminis (2012) consider partitioning a road network into a small number of uniformly congested subnetworks (bathtubs). This paper considers just a single bathtub.